

Algoritmi japanin kirjoitusmerkkien opiskelujärjestyksen optimointiin

Aino Haavisto

Pro gradu -tutkielma

Kielten maisteriohjelma, Aasian kielten opintosuunta

Humanistinen tiedekunta

Helsingin yliopisto

Maaliskuu 2019



Tiedekunta – Fakultet – Faculty Humanistinen tiedekunta		Koulutusohjelma – Utbildningsprogram – Degree Programme Kielten maisteriohjelma	
Opintosuunta – Studieriktning – Study Track Aasian kielten opintosuunta			
Tekijä – Författare – Author Aino Haavisto			
Työn nimi – Arbetets titel – Title Algoritmi japanin kirjoitusmerkkien opiskelujärjestyksen optimointiin			
Työn laji – Arbetets art – Level Pro gradu -tutkielma		Aika – Datum – Month and year Maaliskuu 2019	Sivumäärä– Sidoantal – Number of pages 52 s. + liitteet 10 s.
Tiivistelmä – Referat – Abstract			
<p>Japanin kielen kirjoitusjärjestelmä on eräs maailman monimutkaisimmista, ja sen oppiminen on haastavaa kaikille japania vieraana kielenä opiskeleville. Japania kirjoitetaan yhdistelemällä kahta foneettista merkistöä (hiragana ja katakana) ja morfografista kanjimerkistöä, jossa merkit kuvaavat morfeemeja eivätkä ainoastaan äänteitä. Erityisesti näiden kiinalaisperäisten kanjimerkkien, joita on Japanissa yleisesti käytössä 2136, opiskelu on työlästä. Tässä tutkielmassa on kehitetty japani vieraana kielenä -opiskelijoille uusi, mahdollisimman tehokas ja helpoksi omaksuttava kanjimerkkien opiskelujärjestys ja vertailtu sitä aiempiin opiskelujärjestyksiin.</p> <p>Tutkielmassa on sovellettu Loachin ja Wangin (2016) kiinan kirjoitusmerkkien opiskelujärjestystä optimoivaa laskennallista algoritmia japanin kielelle. Algoritmi järjestää merkit painottaen niiden yleisyyttä ja muodon yksinkertaisuutta, jotta alusta alkaen kielenoppija oppii hyödyllisiä merkkejä. Tämä ensimmäinen vaihe toteutettiin olemassa olevaa korpusa ja merkkitietokantaa käyttämällä. Toisessa vaiheessa algoritmi tekee järjestyksestä vielä hierarkkisen eli nostaa kunkin merkin komponentit opiskeltavaksi ennen komponenteista koostuvaa yhdistelmämerkkiä. Myös tämä vaihe toteutettiin hyödyntäen valmista, joskin työtä varten muokattua tietokantaa.</p> <p>Tutkimuksessa selvisi algoritmin luomaa järjestystä tarkastellessa, että Loachin ja Wangin algoritmin tuottamaa järjestystä on pidettävä vasta raakaversiona opiskelujärjestyksestä. Erityisesti kanjeihin liittyvien temaattisten kategorioiden huomiotta jättäminen tekee järjestyksestä joiltain osin vaikeasti hahmotettavan. Tämä vuoksi tutkielmassa pohditaan erilaisia mahdollisuuksia kehittää algoritmin tulosta edelleen ja esitellään noin 250 ensimmäiselle merkille algoritmiin perustuva, mutta käsin muokattu järjestys. Luotu järjestys noudattaa edelleen periaatteita yleisten merkkien opettamisesta hierarkkisessa järjestyksessä, mutta hyödyntää lisäksi temaattisia kategorioita merkkien järjestämisessä.</p> <p>Opiskelujärjestyksen luomisen lisäksi työssä tutkittiin tätä järjestystä hyödyntävän oppimateriaalin luomista puoliautomaattisesti korpuksia ja sanakirjatietokantoja hyödyntämällä. Merkeistä luotiin internetiin lista, johon on lisätty kunkin merkin kohdalle mm. yleisiä sanoja sisältäviä esimerkkilauseita. Lisäksi luotiin kertaukseen soveltuva virtuaalinen korttipakka. Tässä työssä tuotettu materiaali muodostaa käyttökelpoisen pohjan, mutta automaattisesti luotua oppimateriaalia on vielä kehitettävä ennen kuin siitä saa käytännössä hyödynnettävää materiaalia. Sekä itse opiskelujärjestystä että oppimateriaalia voidaan parantaa sekä ihmisen tekemällä työllä että muokkaamalla käytettyjä tietokoneohjelmia.</p>			
Avainsanat – Nyckelord – Keywords japanin kirjoitusjärjestelmä, kanjimerkit, japanin opetus, laskennallinen kielitiede, korpuslingvistiikka			
Säilytyspaikka – Förvaringställe – Where deposited Keskustakampuksen kirjasto			
Muita tietoja – Övriga uppgifter – Additional information			

Sisällysluettelo

1. Johdanto.....	4
1.1. Translitteraatio ja käytetty termistö.....	5
1.2. Tutkimuksen tarkoitus.....	6
2. Japanin kirjoitusjärjestelmä.....	7
3. Kanjimerkkien oppiminen ja opiskelujärjestyksen merkitys.....	13
3.1. Katsaus kanjimerkkien oppimisen psykolingvistiikkaan.....	13
3.2. Kanjien opiskelustrategioita.....	17
3.3. Erilaisia opiskelujärjestyksiä kanjeille.....	18
Opiskelujärjestysten luokittelua.....	19
Japanilaisten lasten standardoitu kanjiopetus.....	20
JFL-opetuksessa käytettyjä kanjien opiskelujärjestyksiä.....	20
3.4. Kanjiopetuksesta Suomessa.....	22
4. Algoritmiikan käyttö opiskelujärjestyksen luontiin.....	23
4.1. Loachin ja Wangin algoritmi.....	24
4.2. Loachin ja Wangin algoritmin luoman opiskelujärjestyksen edut ja puutteet.....	26
4.3. Muita algoritmeja opiskelujärjestyksen luontiin.....	27
5. Loachin & Wangin algoritmi sovitettuna japanin kieleen.....	27
5.1. Aineisto ja sen käsittely.....	27
5.2. Kanjien jako komponentteihin.....	28
5.3. Algoritmin tuloksen tarkastelua.....	30
5.4. Algoritmin tuloksen muokkaaminen käytettäväksi opiskelujärjestykseksi.....	33
Temaattisten kategorioiden yhdistäminen algoritmin tulokseen.....	34
Yleisten yhdyssanojen huomiointi.....	36
Muut muokkaamisessa käytetyt strategiat.....	37
5.5. Muokattu opiskelujärjestys.....	38
6. Puoliautomaattisesti luotu oppimateriaali.....	40
6.1. Käännösten, ääntöasujen ja esimerkkilauseiden lisääminen puoliautomaattisesti.....	40
6.2. Anki-aikavälikertauspakka.....	43
7. Yhteenveto.....	45
8. Lähteet.....	49
Liitteet.....	53
Liite 1: Algoritmin tuottama opiskelujärjestys.....	53
Liite 2: Muokattu opiskelujärjestys.....	55
Liite 3: Kanjien komponenttilistaan tehdyt muutokset.....	59
Liite 4: Algoritmin ohjelmakoodi.....	61
Liite 5: Kuvaaja opiskeltujen merkkien suhteesta korpuksen yleisimpiin merkkeihin.....	62

1. Johdanto

Japanin kirjoitusjärjestelmää pidetään oikeutetusti yhtenä maailman vaativimmista oppia (Paxton ja Svetanant 2013, 90), ja sen vuoksi japanin kieltäkin pidetään eräänä työläimmistä opiskella (esim. usein viitattu Yhdysvaltojen Foreign Service Instituten lista (FSI's Experience with Language Learning ei pvm.). Funktionaalisen lukutaidon hankkiminen edellyttää satojen merkkien mieleen painamista, ja japania vieraana kielenä opiskelevan on löydettävä itselleen sopivat tehokkaat metodit siihen, mikäli haluaa oppia lukemaan japania. Tämä tutkimus keskittyy etsimään uusia, tehokkaita ja mielekkäitä tapoja oppia japanin kirjoitusmerkkejä.

Japania puhuu äidinkielenään yli 125 miljoonaa ihmistä, ja kieltä opiskelee vieraana kielenä erilaisissa oppilaitoksissa ympäri maailmaa Japan Foundationin mukaan noin 3,6 miljoonaa ihmistä, joista Suomessa tämän arvion mukaan ainakin 1600 (Japan Foundation 2017). Määrä on kasvanut jatkuvasti, joten japanin opiskelu koskettaa koko ajan suurempaa joukkoa ihmisiä. Suuri määrä opiskelijoita sekä Suomessa että maailmalla opiskelee japania itsenäisesti, joten todellista japanin kielestä kiinnostuneiden määrää on mahdoton arvioida.

Oppimateriaalin kasvavaan kysyntään on vastattu niin Suomessa kuin maailmallakin. Suomenkielisiä japanin alkeistason oppikirjoja on julkaistu erityisesti viime vuosina useita, ja internetissä on monia sivustoja, joissa voi opiskella kieltä ilmaiseksi. Lisäksi Worddive-yritys myy suomenkielistä japaninopiskelusovellusta.¹ Suomi-japani-sanakirjoja on useita, ja kaksi kirjoitusmerkkisanakirjaakin löytyy (Bruneau 2009; Tuomi 1998). Japania vieraana kielenä opiskellessa kielen monimutkainen kirjoitusjärjestelmä aiheuttaa oppijalle paljon työtä, ja kirjoitusmerkkien opiskeluun kuluu paljon aikaa. Suomeksi on sekä käännetty (Kardy 2007) että kirjoitettu (Araki-Masonen 2006) japanin tavuaakkosten opiskeluun tarkoitettuja kirjoja, mutta kanjimerkkien opiskeluun suomeksi löytyy ainoastaan yksi sarja, *Kanji de manga* (Kardy ja Hattori 2008), josta on suomennettu neljä ensimmäistä osaa. *Kanji de manga* on kuitenkin käyttömahdollisuuksiltaan hyvin rajattu, ja käytännössä kaikilla japanin kielen kursseilla Suomessa käytetään englanninkielisiä kanjioppikirjoja tai opettajan luomaa omaa oppimateriaalia.

¹<http://www.worddive.com/fi/>

Tässä työssä keskityn nimenomaan kanjimerkkien opiskeluun pyrkimyksenäni luoda uusi tehokas opiskelujärjestys kanjimerkeille ja sitä hyödyntävää suomenkielistä oppimateriaalia. Näin suomenkielisen opiskelijan olisi entistä helpompi aloittaa kanjimerkkien opiskelu ja oppia lukemaan ja kirjoittamaan japania.

Tämän tutkielman ensimmäisessä luvussa kuvaillaan tutkimuksen lähtökohdat, translitteraation konventiot ja määritellään tutkimuskysymys. Luku 2 kertoo japanin kirjoitusjärjestelmästä, ja luku 3 esittelee tämän tutkimuksen teoriataustaa ja olemassa olevia kanjimerkkien opiskelujärjestyksiä. Luvussa 4 kuvaillaan käytetty metodi kirjoitusmerkkien opiskelujärjestyksen algoritmiseen optimointiin, ja luvussa 5 esittelen oman version algoritmistä sekä analysoin ja edelleen kehitän sen tulosta. Tutkielman luku 6 täydentää tutkielmaa esittelemällä luotuun opiskelujärjestykseen liittyvää puoliautomaattisesti luotua oppimateriaalia. Työn päättävässä luvussa 7 tehdään yhteenveto ja pohditaan mahdollisuuksia tuoda luotu opiskelujärjestys käyttöön japanin opetuksessa.

Tämä työ on omistettu Gurulan sohvalle, jolla on ollut elintärkeä rooli tämän tutkielman synnyssä. Lisäksi haluan lämpimästi kiittää ohjaajaani yliopistonlehtori Riikka Länsisalmia sekä kaikkia muita, joiden kanssa olen saanut keskustella työstäni ja joilta olen saanut arvokasta palautetta.

1.1. Translitteraatio ja käytetty termistö

Japania translitteroidaan tässä työssä Hepburn-romanisointijärjestelmää käyttäen lukuun ottamatta pitkien vokaalien a, i, u ja o merkintää, jotka kirjoitetaan kahdella vokaalilla: 書道 *shodoo* 'kalligrafia'. Kanjimerkkien *on'yomi*-lukutapaa merkitään suuraakkosin ja *kun'yomi*-lukutapaa pienaakkosin, milloin ne on tarpeellista erottaa, esim. 道 *michi*, *DOO* 'tie' (ks. tarkemmin lukutavoista luku 2). Merkintä 食べる *ta(beru)* 'syödä', tarkoittaa, että sanan suluissa oleva osa kirjoitetaan kanjimerkkiin liittyvällä hiragana-tavukirjoituksella (nk. *okurigana*). Annan esimerkkeinä käsittelemistäni kanjeista yleensä jonkinlaisen käännöksen suomeksi, vaikka pelkillä kanjimerkeillä ei suoranaisesti olekaan mitään tiettyä merkitystä eikä siten käännöstäkään. Kiinalaisperäiseen, Itä-Aasialle yhteiseen merkistöön viitataan jatkuvasti japanilaisella nimellä *kanji*, ellei asiayhteydessä ole välttämätöntä erotella puhutaanko merkeistä kiinalaisessa vai japanilaisessa kontekstissa.

Käytän tässä työssä japani vieraana kielenä -opiskelusta ja opiskelijoista lyhennettä JFL

(*Japanese as a Foreign Language*). Lyhenne on laajasti käytössä sekä englannin- että japaninkielisessä tutkimuksessa, eikä ota kantaa kielen opiskelupaikkaan (Japanissa/ulkomailla) tai kielten opiskelujärjestykseen, kuten esimerkiksi JSL (*Japanese as a Second Language*) tai termi L2.

Kanjeihin liittyvä termistö ei ole suomeksi kovinkaan vakiintunutta. Käytän kaikista kanjien osista termiä *komponentti* (esim. merkki 想 koostuu komponenteista 相 ja 心, joista ensin mainittu komponentti koostuu vielä komponenteista 木 ja 目). Termi ei tee eroa sen välillä, minkä tason jaosta on kysymys, sillä luomani opiskelujärjestyksen kannalta sillä ei ole väliä. Tärkeää erityistapausta merkkien osista (ks. luku 2), *radikaaleja* (部首 *bushu*) käsittelen vain lyhyesti. *Radikaali* on monimerkityksellinen ja siksi hieman huono termi, mutta suomeksikin englannin kielen vaikutuksesta jo sen verran yleinen, että käytän sitä. Kuitenkin esimerkiksi ensimmäinen laaja suomalainen kanjisanakirja *Suuri Kanjikirja* (Bruneau 2009) käyttää termiä *tunniste* mainiten vaihtoehtoisiksi termeiksi ”radikaali” ja ”juuriosa” (Bruneau 2009, 17).

Komponentti-termiä vastaa kenties lähimmin japanin 要素 *yooso*, jota näkee toisinaan käytettävän. Englanniksi eräs käytetty ilmaus termin *component* ohella on primitiivi (*primitive*), jota käyttävät esimerkiksi Heisig (1977) ja uudemmassa aineistossa Loach ja Wang (2016).

Myös termiä grafeemi (*grapheme*) samassa merkityksessä 'kanjin komponentti' käyttävät esimerkiksi Toyoda (2000, siteerattu teoksessa Rose 2013, 982) sekä Vorobev ja Vorobeva (2012). Termi on sinänsä täsmällinen: esimerkiksi Tieteen termipankki antaa (erääksi) määritelmäksi grafeemille sen toimivan ”pienimpänä merkitystä erottavana yksikkönä kirjoituksessa” (Kielitiede:grafeemi – Tieteen termipankki), mutta koska grafeemi-termiä käytetään harvoin muun kuin foneettisen kirjoituksen yhteydessä, pidän parempana pysyä vakiintuneemmissa ilmauksissa.

1.2. Tutkimuksen tarkoitus

Tämän opinnäytetyön tarkoituksena on tarkastella ja vertailla aiemmin esitettyjä japanin kanjimerkkien opiskelujärjestyksiä, ja muodostaa uusi hierarkkinen opiskelujärjestys kanjeille perustuen Loachin ja Wangin (2016) kiinan hanzi-merkkien opiskelujärjestystä optimoivaan algoritmiin. Kyseessä on eräs ensimmäisistä laskennallisista menetelmistä, joka on kehitetty

optimoimaan kirjoitusmerkkien opiskelujärjestystä. Arvioin tuon tietojenkäsittelytieteilijöiden luoman algoritmin toimivuutta kielen oppijan kannalta ja tutkin erilaisia tapoja optimoida algoritmin tuottamaa opiskelujärjestystä entisestään.

On selvää, että kirjoitusmerkkien opiskelujärjestyksellä on oppijalle väliä. Silti aiheeseen liittyvää tutkimusta on tehty vähän ja japania vieraana kielenä -opetuksessa merkkien opiskelujärjestys perustuu pitkälti oppimateriaalin tuottajien subjektiivisiin näkemyksiin. Kaikki opiskelujärjestykset perustuvat kahteen optimiin, joiden keskinäinen painoarvo riippuu materiaalista: 1) kannattaa opiskella mahdollisimman hyödyllisiä merkkejä; 2) kannattaa opiskella merkkejä mahdollisimman helposti (helposti opittavassa järjestyksessä). Loachin ja Wangin algoritmi käyttää parametreinaan sekä merkkien yleisyyttä että komponentteja ja tarjoaa siten optimoidun kompromissin.

Työssäni keskityn japania vieraana kielenä opiskelevien opiskelijoiden tarpeisiin. Luotu opiskelujärjestys on tarkoitettu ei-kanjitaustan opiskelijoille (非漢字系, *hikanjikei*), jotka eivät siis ennalta tunne kanjeja äidinkiellensä kirjoitusjärjestelmästä (kuten kiinalaiset tuntevat). Eroja kanjien prosessoinnissa kanjitaustaisten ja kanjeja ennalta tuntemattomien opiskelijoiden välillä on tutkittu paljon (Matsumoto 2013), ja eri taustan omaavien opiskelijoiden lukustrategioiden on nähty eroavan toisistaan (Matsumoto 2013). Ei-kanjitaustan opiskelijoille on erityisen tärkeä luoda vahva ymmärrys japanin kirjoitusjärjestelmästä ja sen rakenneosasisista. Myös sen vuoksi komponenttien hyödyntäminen opiskelujärjestyksen luomisessa on järkevää.

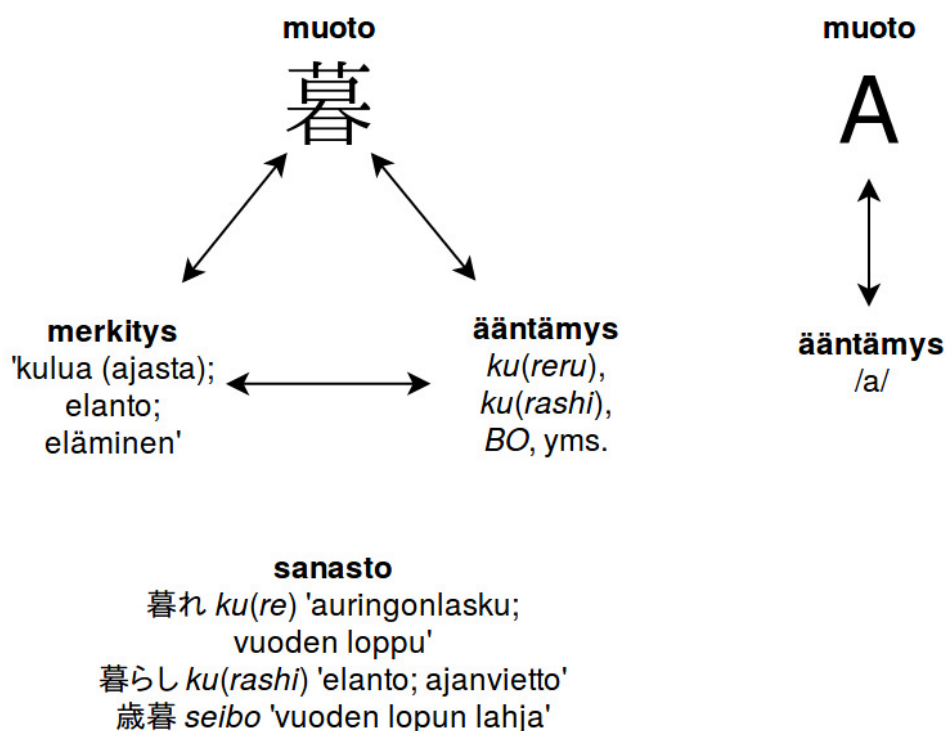
2. Japanin kirjoitusjärjestelmä

Japanin nykyinen kirjoitusjärjestelmä on maailman monimutkaisimpia, ja aiheuttaa siten paljon työtä kielen opiskelijalle. Japania kirjoitetaan kolmea eri merkistöä yhdistelevällä ortografialla, jota nykyään täydentävät myös arabialaiset numerot ja latinalaiset aakkoset. Kolme merkistöä ovat hiragana, katakana ja kanjimerkistö, joista viimeksi mainitun oppimiseen tämä työ keskittyy. Yksinkertaistaen voidaan todeta, että nykyjapanissa kanjeilla kirjoitetaan sisältösanat ja hiraganalla funktiosanat, ja katakanaa käytetään täydentävänä merkistönä esimerkiksi muille kuin kiinalaisille lainasanoille (Iwasaki 2013, 20).

Kirjoitusjärjestelmän yksinkertaistamista on ehdotettu aina silloin tällöin, mutta käytännössä japanilaiset ovat niin sitoutuneita nykyiseen järjestelmään, että lähitulevaisuudessa ei ole

nähtävissä muutosta kanjien asemassa (Carroll 2001, 169–75). Siten ne tulevat olemaan tärkeä osa kielen oppijan urakkaa myös tulevaisuudessa.

Hiragana ja katakana ovat tavumerkistöjä,² joilla voidaan molemmilla ilmaista japanin kielen kaikki foneemit eli kielessä merkitystä erottavat äänteet. Kanjit puolestaan ovat morfografinen merkistö, jonka merkit eivät ilmaise niinkään äänteitä vaan morfeemeja. Toisin kuin foneettisiin merkistöihin, kuten latinalaisiin aakkosiin, kanjeihin liittyy siis aina ääntöasun lisäksi merkitys (ks. kuva 1). Yhdessä nämä kolme merkistöä muodostavat japanin kirjoitusjärjestelmän.



Kuva 1: Morfografisessa merkistössä yhdessä merkissä merkin muotoon yhdistyvät sekä merkitys että ääntämys, fonogrammisessa ainostaan ääntämys. Käytännössä kanjit toimivat aina osana sanastoa, usein useamman merkin kokonaisuutena, ja saavat varsinaisen merkityksensä vasta silloin.

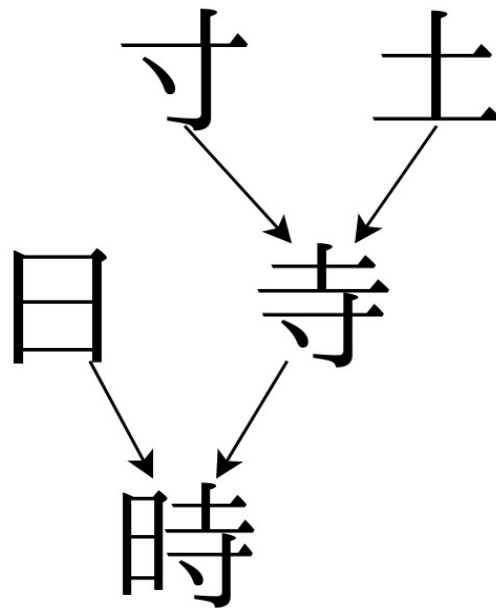
Kanjimerkistö on lainattu alun perin mantereelta yli 1500 vuotta sitten, Kiinasta tai vielä todennäköisemmin kiinaa kirjoitettuna kielenään käyttäneestä Koreasta (Seeley 1991, 4–6).

Kiinassa kanjimerkkejä (kiinaksi *hánzî*) käytetään lähes puhtaasti morfografisesti: yksi merkki

² Koska hiragana ja katakana eivät varsinaisesti kuvaa tavuja vaan moria, vielä tarkempaa olisi puhua niistä mora-merkistöinä. Esimerkiksi はい /hai/ 'kyllä' kirjoitetaan kahdella hiraganamerkillä, sillä se koostuu kahdesta morasta, vaikka tavuja onkin vain yksi.

kuvaa yhden morfeemin merkityksen ja ääntämisen. Japanissa tilanne ei ole näin yksinkertainen, kuten kohta esittelen. Kanji- tai hanzi-merkeistä puhutaan yleisesti logogrammeina (kreikan *logos*, 'sana'), mutta koska yksittäinen merkki ei kuvaa sanaa vaan morfeemia, on morfogrammi tarkempi ilmaus. Vaikka japaniksi tämäkään ei ole täysin täsmällinen termi, tässä työssä kanjeista puhutaan morfogrammeina.

Kanjit eivät koostu satunnaisista viivoista ja niihin liitetyistä ääntöasuista, vaan sekä visuaalinen, semanttinen että fonologinen muoto noudattelevat monia sääntöjä. Visuaalisella tasolla merkkien pienimpiä rakenneosasia ovat *vedot*, joita perinteisesti lasketaan olevan kahdeksan erilaista. Kanjien *komponentit* muodostuvat näistä vedoista ja kukin kanji yhdestä tai useammasta komponentista. Perinteisesti komponentteja on laskettu olevan 214 (nk. *kangxi-radikaalit*, vuonna 1716 keisari Kangxin aikana kootun sanakirjan mukaan nimettyinä (Taylor ja Taylor 2014, 48)), mutta käytännössä merkeissä esiintyy myös komponentteja, jotka eivät löydy tuosta perinteisestä listasta. Kaikki kymmenet tuhannet kanjimerkit koostuvat silti suhteellisen vähälukuisesta määrästä rakennuspalikoita.



Kuva 2: Merkin 時 'aika; tunti' jakautuminen komponentteihin

Yleinen harhakäsitys on, että kaikki kanjit olisivat ”kuvamerkkejä”, mutta käytännössä vain pieni osa nykyisistä merkeistä on piktografisia eli perustuu kuviin (esim. 木 'puu').

Piktografisten merkkien (1) ohella muut kanjien pääluokat ovat 2) yksinkertaiset ideografiset

merkit, jotka kuvaavat jonkin abstraktin konseptin visuaalisesti (下 'alla', 上 'yllä'), 3) yhdistelmä-ideografiset merkit (休 'lepo': koostuu komponenteista 亻 'ihminen' ja 木 'puu') ja 4) foneettis-semanttiset, joista viimeksi mainittuja on suurin osa kirjoitusmerkeistä. Foneettis-semanttiset merkit koostuvat merkin semanttisesta, suurpiirteistä merkitystä kuvaavasta osasta sekä merkin äännearvoa kuvaavasta foneettisesta osasta. Kullakin semanttisella komponentilla on oma perusmerkityksensä, toisilla läpinäkyvämpi kuin toisilla – esimerkkejä selkeistä kategorioista ovat esim. 水 'vesi' (monissa nesteisiin liittyvissä kanjeissa) ja 艸 'ruoho' (monissa kasveihin liittyvissä kanjeissa).

Nämä kanjien pääluokat perustuvat kiinalaisen Shen Xun vuonna 121 tekemään kiinalaisten merkkien luokitteluun. Xun luokittelussa kategorioita olivat edellä kuvattujen neljän peruskategorian lisäksi myös ”foneettinen laina” sekä *zhuanzhu* (轉注, jap. 転注), joista jälkimmäinen on määritelty niin sekavasti, etteivät nykytutkijat ole yhtä mieltä siitä, mistä oikeastaan on kyse (Taylor ja Taylor 2014, 47). Foneettisella lainalla puolestaan tarkoitetaan sitä, että merkki lainautuu tarkoittamaan jotakin muuta samoin ääntyvää morfeemia: yleisenä esimerkkinä 萬, alunperin kiinassa 'skorpioni' /wan/, ja sittemmin '10 000' /wan/ (nykyään japanissa käytetään eri merkkiä: 万 '10 000') (Taylor ja Taylor 2014, 45–46). Nämä viimeksi mainitut kategoriat ovat molemmat huomattavasti ensimmäistä neljää pienempiä ja vähemmän kirjoitusjärjestelmän rakenteen ytimessä.

On merkille pantavaa, että Xun tunnetuimman teoksen *Shuowen jiezin* nimessä on kyse yksinkertaisten merkkien ja yhdistelmämerkkien analysoinnista (Taylorin ja Taylorin (2014, 45) käännös: ”*Explanations of Simple Characters and Analysis of Composite Characters*”) eli jo otsikkoon on nostettu huomio (yhdistelmä)merkkien rakentumisesta komponenteista. Komponenttiperusteisuuden juuret ulottuvat paljon tätäkin kauemmas, vähintään 300-luvulle eaa, jolloin suuren osan nyky-Kiinan alueesta yhdistänyt Qin-keisari määräsi merkit standardoitaviksi: kaikista merkeistä tuli nyt samankokoisia ja -muotoisia (Taylor ja Taylor 2014, 41), ja samalla komponenteistakin tuli pitkälti nykyisenkaltaisia, standardoituja muotoja.

Komponentin erityistapaus *radikaali* on nähdäkseni yksi hankalimmista kanjeihin liittyvistä termeistä, sillä se määritellään eri lähteissä hyvin eri tavoin. Rose (2013, 982) antaa sille yksinkertaisen ja täsmällisen määritelmän ”kanjin juurikomponentti(*root component*) luokittelua varten”. Tämä määritelmä ei anna radikaalille liian suurta painoarvoa eikä sekoita

sitä komponentteihin, kuten valitettavan monissa oppimateriaaleissa esiintyvät määritelmät. Latinalainen merkistö on helppo järjestää kaikille tutun (vaikkakin sinänsä mielivaltaisen) aakkosjärjestyksen mukaan, mutta kanjien määrän vuoksi niiden kohdalla järjestys ei voi perustua ulkoa opeteltavaan listaan. Siksi kanjit on perinteisesti järjestetty kangxi-radikaalien mukaan: kukin kanji löytyy paperisesta sanakirjasta siihen liittyvän radikaalin kohdalta. Usein radikaali on mielekäs merkin kannalta: tyypillisesti semanttis-foneettisen merkin semanttinen tai foneettinen osa, mutta ei aina. Esimerkiksi merkki 希 *KI* 'toivo; harvinainen' liittyy enintään etäisesti radikaaliinsa 巾 *KIN* 'kangas'.

Kanjien fonologinen arvo perustuu tyypillisesti sekin komponentteihin, sillä foneettis-semanttiset merkit ovat syntyneet siten, että uutta sanaa vastaavalla tavalla äännettävää sanaa kuvaavasta merkistä on luotu uusi merkki lisäämällä siihen eri semanttisesta kategoriasta vihjaava merkitysradikaali. Esimerkiksi merkki 青 'sininen', jonka eräs ääntöasu on *SEI*, on yhdistetty komponenttiin 日, 'aurinko', jolloin on saatu kirkasta säätä tarkoittava merkki 晴 *SEI*. Käytännössä nykyjapanissa nämä yhteydet ovat harvoin näin yksinkertaisia, sillä foneettis-semanttiset yhdistelmämerkit on pääosin lainattu kiinasta, ja merkin tai komponentin ääntämys on usein muuttunut ajan myötä.³

Tyypillisimmillään kanji kuvaa siis sekä morfeemin merkitystä että sen ääntämystä. Kiinassa kanjeilla on yleensä yksi tai enintään muutama ääntöasu, mutta japanissa tilanne on monimutkaisempi. Merkeillä on sekä nk. japanilaisia lukutapoja, *kun'yomi* (訓読み) että nk. kiinalaisia lukutapoja, *on'yomi* (音読み). Kun-lukutapa on tyypillisesti morfeemin japanilainen ääntämys, esimerkiksi merkillä 人 *hito*, 'ihminen'. On-lukutapa puolestaan on lainattu kiinasta ja perustuu siten morfeemin kiinalaiseen ääntämisasuun, merkin 人 tapauksessa niitä ovat *JIN* ja *NIN*. Monesti syynä useisiin on-lukutapoihin on se, että merkille on lainattu useampaan kertaan uusia ääntöasuja, ja kiinan ääntämys on ehtinyt välissä muuttua tai uusia ääntöasuja on lainattu eri alueen kielimuodosta.

1) Kun- ja on-lukutapa merkille 人

その	人	は	日本人	です。
<i>sono</i>	<i>hito</i>	<i>wa</i>	<i>nihon+jin</i>	<i>desu.</i>
se	ihminen	TOP	japani+ihminen	COP

”Se henkilö on japanilainen.”

³ Foneettisista komponenteista löytyy mielenkiintoinen koonti osoitteesta: <https://namakajiri.net/nikki/testing-the-power-of-phonetic-components-in-japanese-kanji/#results>

Yleensä kanjimerkkien esiintyessä yksinään, yksittäisinä sanoina, käytetään niiden kun-lukutapaa. Sen sijaan yhdyssanoissa on tavallisesti käytössä on-lukutapa. (Iwasaki 2013, 22). Kaikilla sanoilla ei kuitenkaan ole molempia lukutapoja, ja variaatiota esiintyy paljon. Vaikka kanjien perusluonne on, että ne sisältävät sekä merkityksen että äänneasun, niin merkkejä käytetään myös pelkästään jompaakumpaa hyödyntäen. Lukutapaa, jossa kanjista hyödynnetään vain äännearvo sanotaan *ateji*-lukutavaksi. Näitä esiintyy mm. vieraskielisissä erisnimissä, esimerkiksi sanassa 阿弗利加 *afurika* 'Afrikka' merkit eivät tarkoita mitään järkevää yhdessä, vaan ne kirjoittavat ainoastaan mantereen nimen äännearvoa. Tietyllä tapaa *ateji*-lukutavan vastakohta on *jukujikun*, eli lukutapa, jossa sana koostuu merkitykseltään järkevistä kanjeista, mutta ääntöasu on arbitraarinen. Eräät *jukujikun*-sanat ovat hyvin yleisiä, kuten 大人 *otona* 'aikuinen': kanjien merkitykset ovat 'iso' ja 'ihminen', mutta missään muussa sanassa kyseisiä merkkejä ei äännetä tuolla tavalla.

Jukujikun-lukutavat perustuvat usein siihen, että puhutussa japanissa käytössä olleelle sanalle on tarvittu kirjoitustapa, eikä kiinasta ole löytynyt vastaavaa yhden merkin ilmausta tai se ei jostakin muusta syystä ole päätynyt käyttöön. *Ateji* ja *jukujikun* ovat osoituksia siitä, että kiinan morfografisen kirjoitusjärjestelmän lainaamisen typologisesti erilaiseen japaniin on muuttanut järjestelmän eräänlaiseksi sekamuodoksi. Mainittavaa on myös, että sen lisäksi että yhden morfeemin kuvaamiseen voidaan tarvita useampi merkki, kuten 大人 yllä, niin japanista löytyy myös useammasta morfeemista koostuvia sanoja, jotka kirjoitetaan vain yhdellä merkillä. Esimerkiksi 湖 *mizuumi* 'järvi', on yhdyssana, joka koostuu kahdesta sanasta *mizu* 'vesi' ja *umi* 'meri'.

Moninaiset ääntöasut muodostavat haasteen opiskelijalle, mutta ennen kaikkea haasteena on kanjien määrä. Kanjimerkistö on kanamerkistöistä ja esimerkiksi latinalaisista aakkosista poiketen avoin luokka, eikä sen kokoa ei voi määrittää tarkasti. Japanissa on määritelty niin kutsutut *jooyoo*-kanjit (常用漢字) 'yleisesti käytettävät kanjit', joita on tällä hetkellä virallisessa listassa 2136 (常用漢字表 [jooyoo kanji hyoo] 2010). Niiden lisäksi virallisessa listassa erisnimissä käytettäviä merkkejä on miltei tuhat merkkiä lisää, ja varsinkin erityisalojen kirjallisuudessa sekä vanhoissa teksteissä esiintyy vielä tuhansia merkkejä lisää.

Kuten kaikissa kielissä, valtaosa japanilaisesta tekstistä koostuu suhteellisen pienestä joukosta yleisiä sanoja valtaosan sanastosta ollessa hyvin harvinaista (nk. Zipfin laki). Siten opiskelemalla joukon yleisimmissä sanoissa esiintyviä yleisimpiä kanjimerkkejä opiskelija

pystyy lukemaan jo valtaosan yleiskielisestä tekstistä. Esimerkiksi sanomalehtikorpuksessa 500 yleisintä kanjia kattavat 70 % kaikista teksteissä esiintyvistä kanjeista (hiragana ja katakana mukaan laskettuna 88 % kaikista tekstin merkeistä) (Nozaki ja Shimizu 2000) ja 1000 yleisintä kanjia kattavat jo 95 % kaikista sanomalehdissä esiintyvistä kanjimerkeistä (Nozaki, Yokoyama, Isomoto ja Yoneda 1996).

3. Kanjimerkkien oppiminen ja opiskelujärjestyksen merkitys

Tässä luvussa esitellään kanjien oppimiseen liittyvää aiempaa tutkimusta sekä sitä, millaisia opiskelujärjestyksiä kanjien opiskelussa käytännössä käytetään. Pääpaino on merkkien opiskelujärjestykseen liittyvällä tutkimuksella. Kanjien opiskelujärjestyksestä luotaessa tasapainotellaan kahden osittain ristiriitaisen optimin, hyödyllisyyden ja helppouden, välillä: 1) kannattaa opiskella mahdollisimman hyödyllisiä (yleensä: yleisiä) merkkejä ja 2) kannattaa opiskella mahdollisimman helppoja merkkejä (eli helposti omaksuttavassa järjestyksessä). Kussakin opiskelujärjestyksessä nämä optimit määritellään tarkemmin eri tavoin, ja niille annetaan erilainen painoarvo. Suuri osa tutkimusta keskittyykin enemmän tai vähemmän siihen, millaiset merkit ovat helppoja oppia.

Sekä japanilaiset lapset että japania vieraana kielenä opiskelevat aloittavat kirjoitetun japanin opiskelun hiraganasta ja katakanasta. Jotkin JFL-oppijoille suunnatut oppimateriaalit aloittavat kanjien opetuksen käytännössä samanaikaisesti hiraganan kanssa (esim. Duolingo-sovellus⁴), mutta tyypillisesti kanjien opetus alkaa hieman myöhemmin, kun opiskelijoiden oletetaan jo tuntevan tavuaakkostot.

Japania äidinkielenään opiskelevat lapset opiskelevat kanjit standardijärjestyksessä, jonka on määrittänyt opetuksesta vastaava ministeriö (文部科学省 [*monbukagakushoo*], Opetus-, kulttuuri-, urheilu- ja tiedeministeriö, jatkossa englanninkielisellä lyhenteellä MEXT). Sen sijaan japani vieraana kielenä -opetuksessa kanjeja opetetaan sekä opetusministeriön järjestyksessä että muissa järjestyksissä, joita esittelen luvussa 3.3.

3.1. Katsaus kanjimerkkien oppimisen psykolingvistiikkaan

Ennen kuin esittelen kanjien prosessointiin ja oppimiseen liittyvää tutkimusta, on hyödyllistä määritellä, mitä jonkin *merkin oppimisella* tarkoitetaan. Aakkosten tapauksessa lienee selvä

⁴ <http://duolingo.com>

määritellä, että jonkun aakkosen on oppinut siinä vaiheessa, kun osaa lukea ja kirjoittaa sen. Viimeistään aakkoston hallitsemisessa lienee syytä määritellä myös, että osaa lukea ja kirjoittaa sen mielekkäässä kontekstissa, eikä ainoastaan yksittäisinä merkkeinä.

Kanjen kohdalla tilanne ei ole yhtä selkeä. Kanjeilla on japanissa useimmiten vähintään kaksi eri ääntämystapaa ja joillakin yleisillä merkeillä on jopa kymmeniä eri ääntötapoja, joista osa on kuitenkin hyvin harvinaisia, eikä ole siksi mielekäästä määritellä, että osataksaan kanjin tulisi osata sen kaikki mahdolliset ääntötavat. Toisin kuin Heisig (1977), olen samaa mieltä tutkijoiden enemmistön kanssa siitä, että jokin tai joitakin ääntämyksiä tulisi kuitenkin osata. Sekä fonologinen, merkkien ääntämykseen perustuva, että semanttinen tulkinta ovat läsnä enemmän tai vähemmän yhtä aikaa kanjeja tulkittaessa (Kess ja Miyamoto 2000, 49; Mori 2014, 405–10). Eri tutkimuksissa on tultu hieman erilaisiin johtopäätöksiin, onko jompikumpi tulkintatapa ensisijaisempi, mutta joka tapauksessa on selvää, että voidakseen todellisessa tilanteessa käyttää kanjeja, on kielen oppijan välttämätöntä tuntea merkin visuaalisen muodon lisäksi sen ääntämys kussakin kontekstissa.

Selvää on, että merkin muoto pitää osata, mutta kuinka tarkasti? Kanjeja on helpompi tunnistaa kuin tuottaa itse, ja yleensä ihmiset tunnistavat huomattavasti enemmän merkkejä kuin mitä pystyvät itse tuottamaan käsin kirjoittamalla. Toisaalta tietokoneaika on muuttanut kirjoittamista: nykylaitteilla on helppo kirjoittaa mikä tahansa kanjisana, kunhan tietää, miten se lausutaan.⁵ Tässä työssä kuvattu algoritmi antaa paljon painoa merkin visuaalisen muodon oppimiselle, joten on oletettavaa, että sitä käyttävä opiskelija oppii hahmottamaan merkkien muodon suhteellisen hyvin.

Kanjen opiskelua ja oppimista on tutkittu paljon, ja silti esimerkiksi Matsumoto (2013, 163) sanoo suoraan, ettei tarkalleen tunneta sitä, miten ihminen kanjeja prosessoi. Suoranaisesti merkkien opiskelujärjestystä ei ole juurikaan tutkittu (Paxton ja Svetanant 2013, 92), tai sitä on tutkittu rajatuissa testiolosuhteissa luokkahuoneiden ulkopuolella, kuten esimerkiksi Lu, Kruss, Webb ja Fox (1999). Yleisemmin kanjeihin liittyvällä tutkimuksella on kuitenkin paljon annettavaa hyvää opiskelujärjestystä luotaessa. Kessin ja Miyamoton (2000, 34–35) mukaan japanilainen psykolingvistinen tutkimus keskittyy pitkälti kanjen prosessoinnin

5 Japanilaista tekstiä syötetään tietokoneella yleensä tavanomaisella qwerty-näppäimistöllä syöttämällä tekstiä foneettisesti (ts. tarkemmin: morapohjaisesti). Konella pystyy siten ”kirjoittamaan” kaikki merkit, joiden ääntämyksen tietää ja joiden muodon tunnistaa sillä tarkkuudella, että pystyy erottamaan ne muista samoin ääntyvistä kanjeista tai yhdyssanoista, joita tietokoneen sanakirja ehdottaa annetun foneettisen muodon perusteella.

tutkimiseen, eli siihen millä tavoin ihmiset tuottavat ja tunnistavat kanjeja. Heidän mukaansa paljon huomiota ovat saaneet erityisesti fonologisten ja semanttisten tekijöiden suhde (voiko merkkejä ymmärtää tuntematta niiden ääntämystä) ja millainen asema radikaaleilla on kanjien ja kanjeista koostuvien yhdyssanojen ymmärtämisessä.

Eri kielten kirjoitusjärjestelmiä käsitellään niiden eroista huolimatta samoilla aivoalueilla vasemmalla aivopuoliskolla, mutta kanjeja käyttäviä kirjoitusjärjestelmiä tulkittaessa mukana on lisäksi myös oikean aivopuoliskon alueita (Mori 2014, 406–7). Selvästi kanjeja tulkittaessa käytetään siis joitakin sellaisia kognitiivisia strategioita, joita pelkkiä fonologisia merkkejä käytettäessä ei käytetä.

Monet kanjien omaksumista ja tunnistamista tutkineet tutkijat suosittavat opettamaan kanjeja hyödyntäen kanjien komponentteja (ks. esim. Chikamatsu 2005; Shen ja Ke 2007; Vorobeva ja Vorobev 2017). Nimenomaan komponenttien ja kanjien hierarkkisuu­den hyödyntämistä suosittelevat Lu ym. (1999), sillä se auttoi japania osaamattomia koehenkilöitä painamaan mieleensä kanjien muodon ja englanninnoksen. Todellisia kielenoppijoita tutkineet Shen ja Ke (2007) totesivat tutkiessaan eritasoisten kiinaa vieraana kielenä opiskelevien yhdysvaltalaisten ($n = 236$) radikaalituntemusta, että hyvin radikaaleja hahmottavat opiskelijat etenevät kiinan opinnoissaan enemmän kuin ne, jotka eivät tunne radikaaleja kunnolla. Heidän mukaansa merkkien komponentteihin tulee kiinnittää huomiota, sillä kielenoppijalla on paljon vähemmän opeteltavaa, jos kunkin merkin vetojen sijasta opettelee merkin komponentit. Everson (2011) puolestaan kirjoittaa, että koska kanjien rakenteesta oppiminen edesauttaa sanaston tunnistamisen kehittymistä ja ylipäänsä lukemista, on merkkien rakenteen opettaminen syytä sitoa tiiviisti opetussuunnitelmaan. Morin (2014) mukaan tietoisuus radikaaleista on kanjituntemuksen peruselementti, ja sen sijaan monesti erityisesti ensikielenoppijoiden opetuksessa painotetusta vetojärjestyksen hallinnasta ei ole juurikaan hyötyä.

Komponentit eivät kuitenkaan kerro kaikkea merkistä tai sanoista, joissa merkki esiintyy. Joissakin tutkimuksissa on myös todettu, että erityisesti alkeistasolla kielenoppijat toisinaan luottavat radikaaleihin liikaa (Mori 2014; Kess ja Miyamoto 2000) jolloin itse kokonaisuus saattaa jäädä ymmärtämättä. Siksi opetuksessa tulisi toisaalta varoa keskittymästä liikaa merkkien komponentteihin.

Kanjien opiskelua pidetään yleisesti vaikeana (opettajien käsityksistä ks. esim. Shimizu ja

Green 2002). Vorobeva (2014) luokittelee kanjien oppimiseen liittyvät ongelmat kolmeen luokkaan: itse kanjeihin liittyviin, opetusmetodeihin liittyviin sekä oppijan asenteeseen liittyviin. Niistä vain yksi, kanjit sinänsä, on sellainen, johon oppija ei pysty lainkaan vaikuttamaan. Koettuun vaikeuteen on kuitenkin monia tapoja vaikuttaa.

Yamashitan ja Marun (2000) tekemässä alkeistason japaninoppijoille suunnatussa kyselytutkimuksessa ($n = 86$) selvisi, että opiskelijat pitivät piktogrammeja helpompina oppia kuin muita merkkityyppejä. Myös katakanamerkkien yhdistelmiksi tulkittavissa olevat kanjit (esim. 加 'lisätä', vrt. katakanamerkit カ /ka/ ja ロ /ro/) ja läpinäkyvät semanttiset yhdistelmämerkit (kuten 岩 'kallio', joka koostuu merkeistä 石 'kivi' ja 山 'vuori') koettiin muita kanjeja helpommiksi oppia. Erilaisten kanjeihin liittyvien kategorioiden näkyviksi tekemisestä voi siis olettaa olevan hyötyä kielenoppijoille. Toisessa kyselytutkimuksessa ($n = 47$) puolestaan selvitettiin muun muassa JFL-opiskelijoiden kokemia ongelmia kanjien oppimisessa: neljännes opiskelijoista mainitsi kanjien epäloogisen tai hämäävän opiskelujärjestyksen olevan ongelma (Librenjak, Vučković, ja Dovedan 2012), joten siltäkin kannalta olisi järkevää kehittää kanjien opiskelujärjestystä.

Kanjien opetus aloitetaan useimmiten muodoltaan mahdollisimman yksinkertaisista merkeistä. On kuitenkin saatu tutkimustuloksia siitä, että kanjien monimutkaisuutta ei ole syytä pelätä liikaa. Natiivipuhujat tunnistavat vetomäärältään, eli eräältä monimutkaisuuden akselilta, keskitasoisia kanjeja (vetomäärä on 8–12) nopeammin kuin näitä yksinkertaisempia tai monimutkaisempia merkkejä (Tamaoka ja Kiyama 2013) ja merkkien frekvenssi vaikuttaa prosessointiin enemmän kuin visuaalinen monimutkaisuus. Ainoastaan ei-kanjitaustan opiskelijoilla merkin suuri vetomäärä vaikeuttaa tunnistamista jonkin verran, mutta tämä ero kanjitaustan opiskelijoihin vähenee opintojen edetessä (Tamaoka 2014, 435–37). Merkin ”monimutkaisuuteen” vaikuttaa toisaalta vetomäärän lisäksi myös ainakin merkin symmetrisyys (Kess ja Miyamoto 2000, 42–43), eli vetomäärältään monimutkainenkin merkki voi olla helposti prosessoitava kokonaisuus. Alle kouluikäisten japanilaisten lasten on todettu oppivan yksinkertaisia kanjeja helpommin kuin hiraganaa tai katakanaa (Kess ja Miyamoto 2000, 138–39). Siitä on siis myös hyötyä, että kanjit ovat suhteellisen monimutkaisia sekä visuaaliselta ulkoasultaan että rakenteeltaan, koska merkeissä yhdistyy fonologinen ja semanttinen aspekti.

Ensikielen ja vieraan kielen oppiminen tapahtuu monilta osin hyvin erilaisten mekanismien perusteella, mutta kanjiopetuksessa on kaksi selvästi toisistaan eroavaa koulukuntaa siinä, pitäisikö opetuksen seurata nykyistä enemmän vai vähemmän natiiviopiskelijoiden mallia. Paxton ja Svetanant (2013) edustavat koulukuntaa, jonka mukaan opetuksessa noudatetaan liikaa japanilaisten lasten opiskelun käytänteitä sekä opiskelujärjestyksen että metodien osalta (Paxton ja Svetanant 2013, 91–92).

Toisaalta Richmond (2005) uskoo, että kanjiopetuksen puutteita tulisi korjata nimenomaan lähentämällä sitä japanilaisten kouluopetuksen kanssa. Richmond korostaa, ettei ensikielen ja vieraan kielen oppiminen sittenkään eroa toisistaan niin paljon, että nykyisenkaltaiset erot kanjien opetuksessa olisivat perusteltavissa. Hänen mukaansa vieraan kielen opetuksessa hylätään suotta monet japanilaisten metodit ja strategiat (Richmond 2005, 57), ja tällä tavalla hän implikoi, että vieraan kielen opetuksen tulisi lähentyä äidinkielen opetusta. Kanjien opiskelujärjestykseen Richmond ei ota suoraa kantaa, vaan ainoastaan mainitsee olettavansa sen olevan merkittävässä asemassa (Richmond 2005, 51).

3.2. Kanjien opiskelustrategioita

Kanjeja voi opiskella lukuisilla eri tavoilla, eivätkä tutkijat tai opiskelijat ole yhtä mieltä siitä, mikä olisi paras opiskelumetodi. Usein kanjien kohdalla mainittuja oppimisstrategioiden kategorioita ovat mekaaninen toisto, muistisäännöt ja kontekstin käyttäminen (esim. Paxton ja Svetanant 2013, 95–97). Nykyään korostetaan ennen kaikkea opiskelutapojen ja -strategioiden monipuolisuutta (mm. Matsumoto 2013, 172; Gamage 2003, 21). Tämän vuoksi myös opiskelujärjestys on järkevää valita siten, että kanjeja pystyy opiskelemaan mahdollisimman monipuolisesti erilaisia strategioita hyödyntäen. Erityisesti kontekstin käyttäminen kanjien opiskelussa on hyvin vaikeaa, mikäli opiskelujärjestys ei lainkaan huomioi merkkien ja niiden muodostamien sanojen kontekstia.

Mekaaninen toisto on useiden tutkimusten mukaan yleisimmin käytetty opiskelustrategia (mm. Gamage 2003; Shimizu ja Green 2002), vaikka sen hyötyjä pidetään yleisesti varsin pieninä. Toisaalta Gamage (2003, 20–21) huomauttaa kyselytutkimuksensa perusteella, että JFL-opiskelijat pitävät itse käyttämiään opiskelustrategioita, erityisesti siis eniten käytettyä mekaanista toistoa, hyödyllisinä kanjien oppimisen kannalta. Samoin Morin, Omorin ja Saton (2016) tutkimuksessa tietotekniikan hyödyntämisestä kanjiopetuksessa suosituin metodi JFL-opiskelijoiden keskuudessa oli perinteiseen toistoon perustuvan virtuaalisen korttipakan

käyttö. Yksinkertaisella toistolla on siis selvästi sijansa yhä kanjiopetuksessa vähintäänkin siksi, että strategia on pidetty opiskelijoiden keskuudessa.

Muistisääntöihin suhtaudutaan näistä kategorioista selvästi ristiriitaisimmin. Muistisäännöllä tarkoitetaan jonkinlaisen tarinan yhdistämistä merkkiin, oli kyse sitten aidosta etymologiasta tai esimerkiksi merkin ulkomuodon opiskelijassa herättämiin mielikuviin. Esimerkiksi merkin 忘 'unohtaa' voisi painaa mieleensä luomalla muistisäännön merkin komponenteista: ”jos *unohtaa* jotakin tärkeää, *sydän* (心) on *kuollut* (亡)”. Rose (2013) tiivistää aiemman tutkimuksen ja 12 englanninkieliselle JFL-opiskelijalle tekemänsä vuoden mittaisen haastatteluihin ja kanjitesteihin perustuvan tutkimuksensa pohjalta, että muistisäännöistä on selvästi apua osalle opiskelijoista. Kaikki eivät kuitenkaan niitä halua käyttää tai koe niistä hyötyvänsä, ja toisaalta niihin turvautuminen ei yksinään tuota sujuvaa kirjallisen japanin taitoa, vaan esimerkiksi ääntämys täytyy opetella useimmiten erikseen (Rose 2013, 989).⁶

Oppimisstrategiat ovat koulumuotoisessa opetuksessa tiiviisti yhteyksissä opetusmetodeihin. Shimizun ja Greenin (2002) tutkimuksen pääosa koostui JFL-opettajille suunnatusta kyselystä ($n = 251$), jossa tutkittiin asenteita kanjeja kohtaan sekä käytettyjä opetusmetodeita. Eniten käytetty metodi oli mekaaninen toisto. Opettajat, jotka olivat keskimääräistä enemmän tietoisia kanjeihin liittyvästä traditiosta, pitivät kanjeja tärkeinä, nauttivat kanjen opettamisesta ja uskovat opiskelijoiden pitävän kanjeista, hyödynsivät todennäköisimmin muita strategioita, eli muistia ja kontekstia (Shimizu ja Green 2002, 238–39). Opettajan asenteilla on siis roolinsa siinä, millä tavoin kanjeja opetetaan, ja opettajan positiivinen suhtautuminen kanjeihin näkyy käytettyjen metodien monipuolisuudessa.

3.3. Erilaisia opiskelujärjestyksiä kanjeille

Toistaiseksi kanjen opiskelujärjestykset ovat perustuneet pitkälti oppimateriaalin luojien henkilökohtaiseen näkemykseen. Sitä ei usein myöskään avata oppikirjan käyttäjälle mitenkään. Vaikka kanjen opiskelua on tutkittu, tutkimuksen tulokset eivät tunnu juurikaan siirtyneen oppimateriaaliksi.

Mahdollinen komponentteihin jakaminen ja komponentteihin perustuvan opetusjärjestyksen luominen on sekin tehty käsin sekä 1970-luvulla (Heisig 1977) että vielä tällä vuosikymmenellä (esim. Wanikani-opiskelusovellus). Algoritmisesti japanin

⁶ Uudemmissa muistisääntöihin perustuvista kanjen opiskelutekniikoista ainakin Wanikani-sovellus liittää muistisääntöihin mukaan myös kunkin merkin/sanan ääntämyksen.

opiskelujärjestystä ei ole tietojeni mukaan optimoitu lainkaan, mutta kiinan osalta vastaavaa tutkimusta on tehty (ks. Yan ym. 2013; Loach ja Wang 2016, joita käsitellään tämän tutkielman luvussa 4). Lähinnä tämänkaltaista datalähtöistä (*data driven*) lähestymistapaa on maksullisen Wanikani-sovelluksen opiskelujärjestyksen optimointi: ohjelman järjestys on luotu kehittäjien henkilökohtaisen näkemyksen perusteella, mutta sitä on alettu 2017 optimoida ohjelman käyttäjien opintomenestyksestä kerätyn datan perusteella (Tofugu 2018).

Opiskelujärjestysten luokittelua

Loachin ja Wangin (2016, 3) mukaan merkkejä voi opiskella vain kahdessa mielekkäässä järjestyksessä: joko frekvenssin mukaan tai hierarkkisessa järjestyksessä, eli opiskellen ensin komponentit ja sitten merkit, joita ne muodostavat. Käytännössä oppimateriaalin tekijöiden mielestä mielekkäinä on pidetty monia muitakin järjestyksiä, ja Paxton ja Svetanant (2013) jaottelevat kanjien opiskelujärjestykset kuuteen luokkaan. Luokat ovat

- 1) piktografeihin ja/tai etymologiaan perustuva
- 2) komponentteihin perustuva (vrt. Loach ja Wangin (2016) termi hierarkkinen järjestys)
- 3) frekvenssiin perustuva
- 4) muotoon perustuva
- 5) muotoon perustuva järjestys
- 6) muu järjestys (Paxton ja Svetanant 2013, 97–102).

Käytännössä Paxton ja Svetanant analysoivat nimenomaan sitä, mitkä merkit ensiksi opetetaan, eivätkä ota kantaa siihen, miten aivan alkeistason jälkeen tulisi jatkaa.

Piktografeihin ja etymologiaan perustuvalla järjestyksellä (1) he tarkoittavat järjestystä, jossa aloitetaan opiskelemalla merkkejä, joiden alkuperä ikoneina, piirrettyinä kuvina tarkoitteistaan, on yhä nähtävissä. Vähiten tutkittuja ja hyödynnettyjä näistä järjestämistavoista ovat *muotoon perustuva* (4) (*form based*), joka tarkoittaa ajatusta opiskella visuaalisesti helposti tunnistettavia merkkejä ensimmäisenä, ja *muotoon perustuva järjestys* (5) (*form based order*), jossa opiskellaan ensin merkkejä, jotka koostuvat selkeistä visuaalisista kokonaisuuksista. (Paxton ja Svetanant 2013). Käytännössä suuri osa oppimateriaaleissa esiintyvistä järjestyksistä kuuluu luokkaan (6), muut järjestykset, joten Paxtonin ja Svetanantin jaon mielekkyys on mahdollista kyseenalaistaa.

Japanilaisten lasten standardoitu kanjiopetus

Japanilaiset koululaiset opiskelevat kuuden ensimmäisen kouluvuoden aikana opetusministeriön määrittelemät 1006 *kyooiku*-kanjia (教育漢字, 'koulutuskanjit'). Uusien kanjien opiskelu jatkuu tämänkin jälkeen läpi kaikkien kouluasteiden. *kyooiku*-kanjien opiskelujärjestys määräytyy kullekin kouluvuodelle jaetun kanjilistan mukaan (学年別漢字配当表 [*gakunenbetsu kanji haitoohyoo*]). Vuonna 2020 alakoulussa opiskeltavien merkkien määrä kasvaa kahdellakymmenellä (MEXT 2017, 28), kun mukaan lisätään Japanin prefektuurien nimissä esiintyviä merkkejä, joita alakoulun yhteiskuntaopin opetuksessa tarvitaan.

Japanilaiseen kanjiopetukseen varsin positiivisesti suhtautuvan Richmondin (2005) mukaan *kyooiku*-kanjien järjestys ”kuvaa tasapainoa merkkien yksinkertaisuuden, merkkien keskinäisen erottuvuuden ja frekvenssin suhteen” (Richmond 2005, 51). Useissa artikkeleissa *kyooiku*-kanjien järjestykseen viitataan nimenomaan frekvenssiin perustuvana, näin tekevät esimerkiksi Paxton ja Svetanant (2013, 90, 99) sekä Librenjak, Vučković ja Dovedan (2012, 1473). Väite on hieman hämmentävä, sillä on helppo nähdä ensimmäisten vuosien kanjilistoilta puuttuvan monia yleisiä merkkejä (esim. 彼 *kare* 'hän (maskuliininen)', joka ei kuulu *kyooiku*-kanjeihin lainkaan). Opetusministeriön viralliset alakoulun opetussuunnitelman perusteet eivät ota lainkaan kantaa siihen, miksi ja missä järjestyksessä tietyt kanjit opetetaan – ainoastaan niiden sopiminen muihin luokka-asteen oppisisältöihin mainitaan (MEXT 2017).

Tätä japanilaista standardia kanjien opetusjärjestyksen suhteen japani vieraana kielenä -opetuksessa käyttää tunnetuista oppikirjoista esimerkiksi Henshall (1989). Paxton ja Svetanant (2013, 91–92) huomauttavat myös, että japanilaiset lapset käyttävät *kyooiku*-kanjien opiskeluun kuusi vuotta, mutta [tavoitteellisilla] aikuisilla JFL-opiskelijoilla on harvoin käyttää vuosikausia kanjien opiskeluun: heille tarvitaan selvästi siis tehokkaampi tapa opiskella kanjeja.

JFL-opetuksessa käytettyjä kanjien opiskelujärjestyksiä

Luvun 3 alussa esittämiini kahteen optimiin (hyödyllisyys vs. helppous) jompaankumpaan puhtaasti nojaavia järjestyksiä on kehitetty, vaikka yleisimpiä ovatkin eri tavoin niitä tasapainottavat sekamuodot. Pelkkään frekvenssiin perustuvia kanjioppimateriaaleja käytetään varsin vähän, sillä tällainen järjestys on oppijalle hyvin raskas kanjien ilmestyessä niin

muodon, merkityksen kuin visuaalisen monimutkaisuuden kannalta täysin satunnaisessa järjestyksessä. Saatavilla on kyllä useita tämän kaltaisia merkkilistoja sekä painettuina kirjoina että esimerkiksi virtuaalisia kertauskorttipakkoja Anki-sovellukseen.⁷

Helppouden optimia on lähestytty monilla tavoin, ja puhtaimmillaan sen voi nähdä pelkästään komponentteihin nojaavissa opiskelujärjestyksissä. Niistä tunnetuin ja laajimmalle levinnyt lienee James W. Heisigin *Remembering the Kanji* -oppikirjan (1977) järjestys. Heisigin metodi perustuu ajatukseen, että merkit kannattaa opetella aina komponentti kerrallaan, komponentit ennen niiden muodostamia yhdistelmämerkkejä (Heisig 1977). Aina Heisigin esitellessä uuden komponentin opetellaan kaikki merkit, jotka voidaan muodostaa tuon komponentin ja kaikkien aiemmin opiskeltujen komponenttien yhdistelminä. Heisigin järjestys ei huomioi lainkaan merkkien frekvenssejä, ja esimerkiksi ensimmäisen 20 kanjin joukossa opitaan merkki 吾 *ware* ('minä', arkaainen muoto, joka on käytössä nykykielessä vain harvoin), joka ei kuulu lainkaan jooyoo-kanjeihin. Tämän vuoksi Heisigin järjestys ei ole tehokas tapa oppia nopeasti kielen keskeisiä merkkejä, vaan siitä on etua vasta, kun on opiskellut satoja merkkejä, kuten Heisigin kiinan kielelle kehittämän vastaavan järjestyksen perusteella toteavat Loach ja Wang (2016, 12).

Käytännössä monissa oppimateriaaleissa sekoitetaan monia erilaisia opiskelujärjestyksiä, ja hyödyllisyyden ja helppouden optimien rinnalle voisi nostaa myös temaattisuuden optimin. Esimerkiksi laajalti käytetyssä *Basic Kanji Book 1* -kirjassa (Kanoh, Takenaka, Ishii ja Shimizu 1990) esitellään kirjan eri kappaleissa merkit erilaisin perustein: Paxtonin ja Svetanantin jaottelusta käytetään lähes kaikkia eri luokkia. Tämänkaltaisella järjestyksellä Kanoh kollegoineen pyrkii opettamaan kanjeja aiemmin tuntemattomille opiskelijoille kaikki kanjien oppimisessa hyödylliset kategoriat, sekä muotoon ja alkuperään perustuvat että semanttiset jaottelut (Kanoh ym. 1990, Esipuhe; Kanoo 1994).

Toinen laajasti käytetty alkeisoppikirjasarja *Genki* (Banno, Ikeda ja Ohno 2011) opettaa sekin kanjeja pääosin temaattisesti, mutta siinä esiintyvät kategoriat ovat miltei kaikki semanttisia, kuten ”vuodenajat” tai ”ilmansuunnat”. Myös *Minna no nihongo* (Nichiguchi 2001) aloittaa kanjiopetuksen temaattisesti: kappaleissa esitellään esimerkiksi verbejä, vapaa-ajan viettoon liittyviä merkkejä tai määriin liittyviä merkkejä. Valintakriteerinä oppikirjan merkeille on

⁷ <https://apps.ankiweb.net/>. Esimerkkinä frekvenssiperusteisesta kanjikorttipakasta *Japanese Kanji Frequency Ordered*, <https://ankiweb.net/shared/info/1647469208>

mainittu ainoastaan merkkein sisältyminen JLPT-taitotasokokeen⁸ toiseksi alimman tason merkkeihin (Nichiguchi 2001, v–vi).

Tällä hetkellä käytössä olevista opiskelujärjestyksistä suurin osa opettaa kanjeja pitkälti optimoiden muita asioita kuin järjestyksen hierarkkisuutta. Vorobeva (2014, 141) on tutkinut työssään 27 eri oppikirjan kanjien esittelyjärjestystä ja todennut tapausten, joissa yhdistelmämerkit esitellään ennen niiden komponentteja olevan yli puolella oppimateriaaleista yli 20 % ja ylimmillään jopa 43 %. Librenjakin, Vučkovićin ja Dovedanin (2012) JFL-opiskelijoille tekemässä kyselytutkimuksessa ($n = 47$) neljännes opiskelijoista mainitsi kanjien epäloogisen tai hämmentävän opiskelujärjestyksen olevan ongelma, joten siltäkin kannalta olisi järkevää kehittää kanjien opiskelujärjestystä.

3.4. Kanjiopetuksesta Suomessa

Suomalaista kanjioppikirjaa ei ole kirjoitettu, vaikka sanakirjoja ja japanin oppikirjoja löytyy. Suomalaisista oppikirjoista mm. Takako Karppisen *Japanin kielen alkeet* (Karppinen 1999) ja Junichiro Okuran *Elävää japania 1* (Okura 2011) opettavat kanjimerkkejä sanaston ja kieliopin ohella, mutta niidenkin painopiste on muualla kuin kirjoitusmerkeissä. Kanjeja opiskellaan Suomessa laajasti englanninkielisen *Basic Kanji Book* -oppikirjan (Kanoh ym. 1990) avulla (mm. Helsingin yliopistossa sekä useimmilla pääkaupunkiseudun kansalaisopistojen kursseilla (Ilmonet 2018)).

Japanin kieltä opetetaan Suomessa nykyään myös joissakin lukioissa ja peruskouluissa. Japanille lukiossa alkavana vapaaehtoisena B3-kielenä on ollut vuodesta 2016 virallinen opetussuunnitelma yleisotsikolla *Aasian ja Afrikan kielet* (Opetushallitus 2015, 118–120), ja yläkoulussa alkavan B2-kielen opetussuunnitelmaa valmistellaan parhaillaan. Opetushallituksen virallinen opetussuunnitelma on hyvin lyhyt ja opetuksen tavoitteissa sanotaan ainoastaan, että tavoitteena on että oppija ”kehittyy kohdekielen kirjoitusmerkistön hallinnassa” ja saavuttaa ”[eurooppalaisen kielitaidon viitekehyksen tason] A1.3 kirjoitetun tekstin ymmärtämisen ja tuottamisen taidoissa”. Kirjoitusmerkit mainitaan lähes jokaisen valtakunnallisen syventävän kurssin kohdalla, mutta mitään tarkkoja tavoitteita ei luoda, eikä niitä niin erilaisille kielille kuin johdannossa mainitut japani, kiina ja arabia voisikaan järkevästi luoda.

⁸ JLPT eli Japanese Language Proficiency Test on kansainvälinen japanin taitotasokoe, jonka järjestää vuosittain eri puolilla maailmaa, myös Suomessa, Japan Foundation.

Japanin opetusta suomalaisissa kouluissa edistävä Ippo-hanke⁹ on luonut japanille oman tarkemman opetussuunnitelman (Ippo-hanke 2016), samoin on tehnyt vastaava kiinan kielen opetusta kouluissa edistävä Yanzu-hanke kiinan kielelle (Yanzu-hanke 2012). Yanzu-hankkeen opetussuunnitelma on merkkien opetuksen suhteen lähes yhtä avoin kuin virallinen eri kielille yhteinen opetussuunnitelma, mutta Ippo-hankkeen opetussuunnitelma kuvailee kanjiopetuksen tavoitteita hieman tarkemmin.

Ippo-hankkeen kanjiopetussuunnitelma on kuitenkin jossakin määrin ristiriitainen.

Valtakunnalliset syventävät kurssit -otsikon alla sanotaan "[l]ukion kahdeksan kurssin aikana opiskellaan Japanissa alakoulun ensimmäisellä ja toisella luokalla opeteltavat n. 250 perus kanji-merkkiä", mutta kurssikohtaisten kuvausten kanjeja käsittelevät osuudet eivät vastaa noita ensimmäisiä kyooiku-kanjeja. Kanjeja aletaan opetussuunnitelman mukaan varsinaisesti opiskella kurssilla 3, ja kursseilla 4–8 opiskellaan merkkejä temaattisesti, esimerkiksi kurssilla 5 *Hyvinvointi ja huolenpito* "Opetellaan arkipäiväiseen elämään, hyvinvointiin, ihmissuhteisiin ja elämänvaiheisiin liittyviä kanji-merkkejä." ja kurssilla 8 *Yhteinen maapallomme* "Opetellaan luontoon, asuinympäristöön ja kestävään elämäntapaan liittyviä kanji-merkkejä". Erityisesti loppupään kurssien teemat menevät huonosti yhteen japanilaislapsille opetettavien ensimmäisten merkkien kanssa, sillä lapsille opetetaan pitkälti hyvin konkreettisten käsitteiden merkkejä sekä alkuopetuksen opetussisältöihin liittyviä merkkejä.

Käytännössä japania ei ole suomalaisissa lukioissa opetettu vielä koskaan kaikkia kahdeksaa, ja vain muutamassa koulussa useampi kuin muutama (Ulla Saari, suullinen tiedonanto 9.9.2018). Koska tarkkaa, ristiriidatonta ja käytössä testattua opetussuunnitelmaa japanin kirjoituksen opiskeluun ei Suomessa ole, on meillä tilaa keskustella kanjiopetuksen kehittämisestä. Tämä tutkielmani on oma panokseni keskusteluun siitä, mitä merkkejä ja millaisessa järjestyksessä voisi olla hyödyllistä opettaa. Merkkien opetusjärjestyksen pohdinnasta hyötyvät niin erilaisissa oppilaitoksissa kieltä opiskelevat kuin itseopiskelijatkin.

4. Algoritmiikan käyttö opiskelujärjestyksen luontiin

Kuten edellä totesin, japanin merkkien opiskelujärjestykset perustuvat harvoin tutkimukseen

⁹ Ippo – Japani peruskoulussa ja lukiossa (2012–2018) oli Opetushallituksen rahoittama valtakunnallinen hanke, jossa pyrittiin kartoittamaan ja kehittämään japanin opetusta suomalaisissa peruskouluissa ja lukioissa. 2018 alkanut *Aasian ja Afrikan kielten hanke* jatkaa osaltaan sen työtä (ks. hankkeen sivusto <https://aasianjaafrikankielet.wordpress.com/>).

vaan luojansa intuitioon. Nykyaikana on kuitenkin mahdollista tietokoneavusteisesti luoda opiskelujärjestyksiä, jotka perustuvat tarkasti haluttuihin parametreihin, olivat parametrit sitten merkkien yleisyys, niiden komponenttirakenne, etymologia tai jokin muu. Tämän kaltaista tutkimusta ei ole tehty vielä paljon, mutta joitakin esimerkkejä on olemassa.

4.1. Loachin ja Wangin algoritmi

Loach ja Wang (2016) pohjaavat kiinan hanzi-merkkien opiskelujärjestyksestä optimoivan algoritminsa Yan ym. (2013) esittelemään algoritmiin, mutta osoittavat oman versionsa olevan Yan ym. algoritmia tehokkaampi. Algoritmi muodostaa merkeille topologisen järjestyksen, joka optimoi merkkien *keskeisyyden* (*centrality*). Topologisella järjestyksellä tarkoitetaan järjestystä, jossa kunkin järjestettävän alkion riippuvuudet tulevat ennen alkioita itseään. Hanzi- tai kanjimerkkien tapauksessa siis merkin kaikki komponentit (ja niiden alakomponentit) tulevat ennen itse merkkiä. Esimerkiksi 火 'tuli' ja 田 'riisipeltö' tulevat topologisessa järjestyksessä ennen merkkiä 畑 'viljelty pelto'. Luodun järjestyksen on siis tarkoitus luoda laskennallisesti tasapaino opiskelujärjestyksen helppouden ja hyödyllisyyden optimien välille.

Keskeisyys, kuten muutkin algoritmin parametrit, ovat tutkijoiden itsensä määrittelemiä: keskeisyys n on merkin frekvenssi korpuksessa jaettuna *oppimisen hinnalla* (*learning cost*), $n = f/c$. Hinta c pyrkii parametrisoimaan työmäärän, joka kielenoppijalta kuluu kunkin merkin muodon oppimiseen: se muodostuu *primitiivien* (Loachin ja Wangin termi merkkien pienimmille jakamattomille osille) kohdalla kaavalla $c = 1 + y * s$, jossa s on merkin vetojen määrä ja y jokin vakio, jonka Loach ja Wang (2016) asettavat arvoon 0,1 perustelematta valintaansa. Useammasta primitiivistä koostuvan merkin hinta puolestaan on sen muodostamiseen tarvittavien kombinaatioiden (plusmerkkien) määrä m , esimerkiksi 的 muodostuu yhdestä kombinaatiosta (白 + 勺) ja 茶 kahdesta (艹 + 人 + 木). (Loach ja Wang 2016).

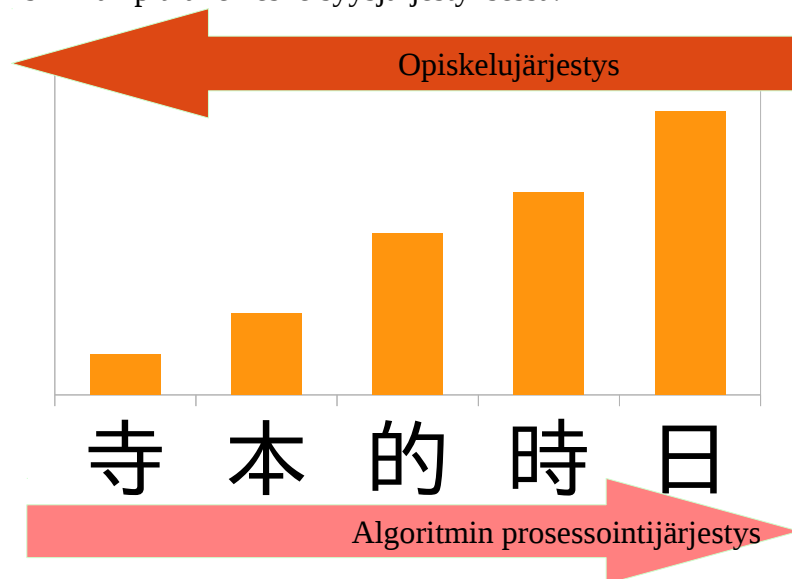
Oppimisen hinnat edellä annetuille esimerkeille parametrilla $y = 0,1$ ovat siis 白 = 1,5; 勺 = 1,3; 的 = 1; 艹 = 1,3; 人 = 1,2; 木 = 1,4 ja 茶 = 2. Kun kullekin merkille on laskettu sen rakenteen perusteella oppimisen hinta ja esiintymäfrekvenssin avulla keskeisyys, voidaan merkit sijoittaa keskeisyyden mukaan järjestykseen. Tämä järjestys ei ole kuitenkaan vielä optimaalinen artikkelissa tarkoitettulla tavalla, sillä järjestys ei ole vielä topologinen. Topologiseen järjestykseen, joka minimoi kumulatiivisen oppimisen hinnan ("opiskelijan

kokonaisvaivan”), merkit asetetaan seuraavalla algoritmilla:

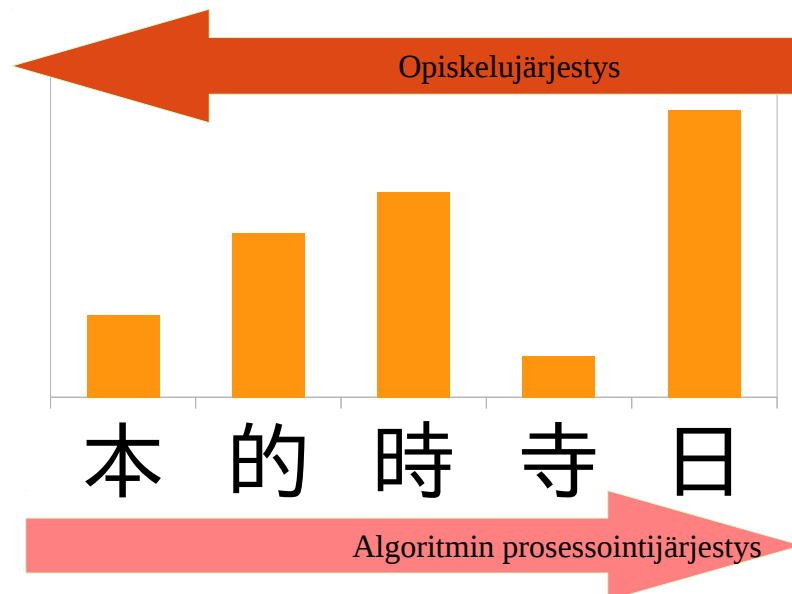
1) Aseta merkit keskeisyyden mukaan järjestykseen pienimmästä suurimpaan. (Kuva 3)

2) Tutki järjestyksessä pienemmästä keskeisyydestä suurempaan jokaisesta merkistä, onko sen kukin komponentti keskeisempi kuin itse merkki (eli tulee opiskelujärjestyksessä ensin). Jos on, niin säilytetään järjestys, mutta jos jokin komponentti on vähemmän keskeinen, siirretään se tutkitun merkin oikealle puolelle eli opiskeltavaksi ennen tuota merkkiä. (Kuva 4)

Algoritmilla aikaansaatu järjestys on siten topologinen (komponentteihin perustuva) järjestys, joka on mahdollisimman pitkälle keskeisyysjärjestyksessä.



Kuva 3: Kanjeja järjestettynä keskeisyyden mukaan (esimerkkikuvan järjestys vain viitteellinen). Huomaa, miten merkki 時 on keskeisempi kuin komponenttinsa 寺.



Kuva 4: Komponentti 寺 siirretty yhdistelmämerkin 時 edelle opiskelujärjestyä.

4.2. Loachin ja Wangin algoritmin luoman opiskelujärjestyksen edut ja puutteet

Yllä kuvattu opiskelujärjestys tarjoaa laskennallisesti luodun kompromissin kirjoitusmerkkien opiskelun molempien optimien, hyödyllisyyden ja helppouden, suhteen. Keskeisyyden mukaan järjestetty lista on lähellä käytetyn korpuksen frekvenssilistaa eli koostuu kielen käyttäjälle hyödyllisistä merkeistä. Se painottuu kuitenkin visuaalisesti yksinkertaisten suuntaan eli Loachin ja Wangin termein merkkeihin, joiden oppimisen hinta on pieni. Vaikka tapoja luoda opiskelujärjestys kanjeille on yhtä monta kuin oppimateriaalin tekijää, niin kaikissa tapauksissa merkkien visuaaliselle yksinkertaisuudelle on annettu ainakin jonkin verran painoa. Jo ennen järjestyksen korjausta hierarkkiseksi mukana on siis helppousaspekti, ja erityisesti korjauksen jälkeen merkit ovat helposti opittavassa järjestyksessä.

Algoritmi ottaa huomioon kuitenkin ainoastaan merkkien muodon niiden oppimiseen kuluva työmäärä arvioidessaan. Merkin muodon lisäksi kielenoppijan pitää oppia merkki myös sanaston osana, eli painaa mieleensä merkin ääntöasu ja merkitys kussakin sanassa. Tätä vaikeammin parametrisoitavaa puolta Loach ja Wang (2016) eivät huomioi. Toinen selkeä algoritmin puute on käytettyjen parametrien suhteellisen mielivaltainen määrittely. Kuinka paljon vaikeampaa tai helpompaa ihmisen on oppia uusi, tietystä määrästä vetoja koostuva komponentti tai jo tuttujen komponenttien yhdistelmä? Miten oppimisen hintaa eli vaikeutta oppia muistamaan ja käyttämään kutakin merkkiä voi mitata?

Loachin ja Wangin (2016) työssä ei ole kyse psykolingvistiikan alan tutkimuksesta, kuten ei tässäkään tutkimuksessa. Siksi kysymystä parametrien mielekkyydestä ei käsitellä tässä ihmisen kognition kannalta, vaan ainoastaan algoritmin lopputuloksen mielekkyyden kannalta. Tuloksiin japanin kielen osalta voi tutustua tämän työn seuraavissa pääluvuissa. Aiemmat kanjien opiskelujärjestykset ovat useimmiten perustuneet luojansa intuitioon: tässä työssä kyse on intuition tukemisesta laskennallisella tiedolla kanjien todellisesta esiintyvyydestä sekä niiden rakentumisesta komponenteista.

Algoritmin tuottama opiskelujärjestys vaatii muokkausta ennen kuin siitä tulee käytännössä mielekäs opiskelujärjestys – tätä puolta Loach ja Wang (2016) eivät lainkaan käsittele, mutta itse paneudun juuri siihen seuraavissa luvuissa. Algoritmin tulosta kannattaakin pitää lähtökohtana, jota on helppo muokata monella tavalla haluamaansa suuntaan.

4.3. Muita algoritmeja opiskelujärjestyksen luontiin

Japanin kielen opiskelujärjestyksen optimoinnista löytyy eräs mielenkiintoinen esimerkki. Tao Romera Martinez on optimoinut merkkien ääntöasujen opiskelua hieman vastaavasti kuin Loach ja Wang itse merkkien opiskelua: mahdollisimman yleisesti käytetyt ääntöasut opiskellaan ensin ja esimerkkilauseet tulevat hierarkkisessa järjestyksessä siten, että niissä esiintyy vain aiemmin opittuja sanoja (Romera Martinez 2016). Romera Martinezin työssä oletetaan opiskelijan opiskelleen jo itse merkit Heisigin menetelmällä, joten hänen kehittämänsä Iikanji-sovellus¹⁰ sopii vain tietyille suhteellisen pienelle joukolle japanin opiskelijoita.

Loach ja Wangin perustana oleva Yan kollegoineen (2013) antaa kiinalaisille merkeille erään merkkien rakenteen hierarkkisuutta hyödyntävän järjestyksen, mutta ei välttämättä optimaalisinta. Heidän luomansa järjestys perustuu siihen, että merkit järjestetään suunnatuksi syklittömäksi verkoksi: verkon solmuina toimivat merkit/komponentit ja kukin merkki/komponentti on yhdistetty suoriin alakomponentteihinsa. Oppimisen hinta on määritelty samoin kuin miten Loach ja Wang sen määrittelevät, mutta opiskelujärjestys luodaan painottaen hierarkkisuutta ja merkkien yleisyyttä eri tavoin. Loach ja Wang (2016) osoittavat oman järjestyksensä olevan kyseisillä parametreilla mitaten tehokkaampi, toisin sanoen minimoivan kumulatiivista oppimisen hintaa tehokkaammin kuin edeltäjänsä.

5. Loachin & Wangin algoritmi sovitettuna japanin kieleen

Tässä luvussa esittelen oman panokseni japanin opiskeluun: sovellan edellisessä luvussa esiteltyä algoritmia japanin kielen kirjoitusjärjestelmään, optimoin algoritmin tulosta manuaalisesti ja analysoin tuloksia. Kaikki työhön liittyvä ohjelmakoodi sekä pohjana käytetty data ovat saatavilla internetissä osoitteessa:

<https://github.com/ahaavisto/ahaavisto.github.io/tree/master/gradukoodi>

5.1. Aineisto ja sen käsittely

Rajasin järjestykseen laitettavat kanjit jooyoo-kanjeihin. Niiden 2136 merkin osaaminen on tavoitteena niin oppivelvollisuuden päättävillä japanilaisilla kuin monilla JFL-

¹⁰ <http://benkyo.co/iikanji/>

opiskelijoillakin. Vaikka opiskelija ei jatkaisikaan niin pitkälle, vaan tyytyisi esimerkiksi 500 merkin opiskeluun, on luontevaa, että nuo 500 keskeisintä on valittu jooyoo-kanjeista. Toisaalta samalla logiikalla toimivan opiskelujärjestyksen ulottaminen koko jooyoo-joukkoon kannustaa jatkamaan opintoja siihen saakka, että kaikki jooyoo-merkit ovat hallussa.

Käytin kanjien esiintymäfrekvenssejä *Balanced Corpus of Contemporary Written Japanese* (BCCWJ) -korpuksen luokitelluista alakorpuksista (NINJAL 2011). BCCWJ-korpus pyrkii olemaan mahdollisimman tasapainoinen otos kirjoitetusta nykyjapanista sen kaikissa eri muodoissa (Summary: *Balanced Corpus of Contemporary Written Japanese* 2009), joten se soveltuu hyvin lähteeksi yleisluontoista kielen oppimateriaalia luodessa. Sekä korpuksen että frekvenssilistan siitä on koostanut Japanin National Institute for Japanese Language and Linguistics, NINJAL (国立国語研究所 [*kokuritsu kokugo kenkyuujo*]). Poistin tarpeettomat tiedot ja pelkistin listan tiedostoksi, jossa kullakin rivillä on ainoastaan merkki ja sen frekvenssi koko käytetyssä korpuksessa. Mikäli haluttaisiin optimoida kanjien opiskelua jotakin tiettyä tekstilajia tai tietyn alan tekstejä varten, algoritmin voisi ajaa jotakin BCCWJ-korpuksen alakorpuksen frekvenssilistaa käyttäen, tai käyttää jotakin aivan muuta korpusta.

Kanjien komponenttilistan hain Kioton yliopiston Chise-tutkimusprojektin tietokannasta (CHISE (CHAracter Information Service Environment) project 2018). Käsittelen seuraavassa alaluvussa tarkemmin kanjien jakamista komponentteihin tässä työssä.

Kunkin kanjin vetojen lukumäärän luin Kanjidic2-tietokannasta (Electronic Dictionary Research and Development Group ei pvm.). Käytin samaa tietokantaa myös jooyoo-merkkien tunnistamiseen eli siihen, että rajasin Chise-projektin merkkilistan siten, että se käsittää vain jooyoo-kanjit ja niiden komponentit.

Käytin pohja-aineiston muokkaamiseen ohjelmallisesti luettavaksi, itse algoritmin ohjelmointiin ja sen jälkeen tulosten muokkaamiseen ihmisluettavaan muotoon python3-ohjelmointikieltä. Toteutin algoritmin täsmälleen Loachin ja Wangin (2016) artikkelissaan kuvaamalla tavalla, ja varsinaisen algoritmin ohjelmakoodi on luettavissa liitteessä 4. Suurin osa ohjelmakoodistani tekee pohjatyötä algoritmille käsittelemällä pohja-aineistoa ja laskemalla merkeille algoritmin vaatiman hinnan ja keskeisyys-arvon.

5.2. Kanjien jako komponentteihin

Kanjien jakaminen komponentteihin ei ole yksikäsitteistä. Perinteisistä kangxi-radikaaleista

olla yksimielisiä, mutta monien merkkien kohdalla niidenkin osalta on kyseessä lähinnä sopimus, eikä se, että valittu radikaali olisi kuvaavin millään tavalla: aina radikaali ei ole semanttinen eikä foneettinen, eikä erotu edes visuaalisesti. Esimerkiksi hyvin yleisen merkin 事 'asia' radikaali on 亼 'koukku', vaikka yhtä hyvin se voisi olla 口 'suu' tai 丩 'sian kärsä'.

Modernilla ajalla kun, yleensä muut kuin japanilaiset, kirjailijat ovat pyrkineet luomaan systemaattisempia ja/tai helppokäyttöisempiä komponenttijakoja kanjeille sanakirjoja tai oppikirjoja varten. Esimerkiksi Halpern (1990) kehitti *New Japanese-English Character Dictionary* -sanakirjaansa varten SKIP-nimellä tunnetun komponenttijaon, joka nojautuu vain osittain perinteisiin radikaaleihin. Myös Heisigin (1977) komponenttijako eroaa monesti perinteisestä. Uudemmassa tutkimuksessa Vorobev ja Vorobeva (2012, 47–48) systematisoivat tekemänsä komponenttijaon periaatteet, mutta siitä huolimatta määritelmä on hieman sekava ja nojaa osittain perinteisiin radikaaleihin. Toisessa tutkimuksessaan Vorobeva vertaili yhdeksässä eri komponenttijaossa syntyvien pienimpien, jakamattomien komponenttien määrää, joka vaihteli 117 ja jopa 632:n välillä (Vorobeva 2011, 18).

Käyttämäni Chise-projektin tietokanta jakaa kanjit komponentteihin tarkemmin kuin vain irrottamalla radikaalin muusta merkistä. Noudatin pitkälti heidän luomaansa komponenttijakoa, sillä se oli tunnetun yliopiston tutkijoiden luoma ja lopputuloskin vaikutti mielekkäältä. Jouduin kuitenkin myös muokkaamaan sitä itse. Chisen jako osoittautui tämän projektin kannalta ongelmalliseksi kolmella tavalla: 1) jako perustui liikaa perinteisiin radikaaleihin, 2) kukin merkki tai komponentti koostui lähes aina tasan kahdesta (ala)komponentista ja 3) jako sisälsi runsaasti komponentteja, jotka eivät kuulu merkistöstandardeihin. Näiden puutteiden vuoksi kävin komponenttilistan useita kertoja läpi ja muokkasin käsin niitä kohtia, joissa komponenttijako ei sopinut Loachin ja Wangin algoritmin logiikalle tai johti näkemykseni mukaan muutoin opiskelijaa harhaan. Tarkemmin Chise-projektin komponenttijakoon tekemiäni muutoksia kuvaillaan liitteessä 3.

Ensimmäinen ongelma, eli että tietokanta perustuu pitkälti perinteisiin radikaaleihin, oli odotettavissa. Vaikka kangxi-radikaalit ovat tuttuja kaikille kanjeja (tai hanzveja) opiskelleille, komponentteihin jako niiden perusteella on kaukana systemaattisesta eikä siten vastaa täysin kanjeja ennalta tuntemattoman alkeisopiskelijan intuitiota merkkien osista.

Toinen ongelma eli Chise-projektin tapa jakaa merkit lähes aina tasan kahteen komponenttiin aiheutti enemmän työtä tämän projektin kannalta. Useissa kohdissa monimutkaisen merkin

jako kahteen komponenttiin on turhan mekanistinen, varsinkin siinä vaiheessa kun alkaa tuntea yleisiä komponentteja. Esimerkiksi merkin 数 'numero; usea' komponentit ovat Chisen mukaan 娄 ja 攴, joista oikeanpuolimmainen komponentti 攴 (kangxi-radikaali 66, 'isku') esiintyy kyllä paljon komponenttina, mutta vasemmanpuolimmainen ei jooyoo-kanjien joukossa missään muualla kuin tässä merkissä. Sen sijaan vasemman komponentin alakomponentit, 米 'riisi' ja 女 'nainen' ovat sekä yleisiä komponentteja että esiintyvät myös yksinään merkkeinä. Opiskelijan kannalta on siis paljon havainnollisempaa esitellä merkki 数 koostuvana kolmesta komponentista [米,女,攴] kuin kahdesta komponentista [娄,攴]. Tämänkaltaisia muutoksia tein paljon enemmän kuin edellä kuvatun kaltaisia: 127 kappaletta jooyoo-kanjien ja niiden komponenttien osalta.

Kolmas ongelma johtui siitä, että Chisen käyttämä jako käyttää myös useita sellaisia komponentteja, joita ei löydy Unicodesta tai muista merkistöstandardeista. Tämä on väistämätöntäkin, koska komponenttijaot eivät ole yksikäsitteisiä, eikä kaikkien eri ihmisten ehdottamia eri komponentteja olisi mielekästä lisätä standardiin. Kymmenet tämän kaltaisista standardeihin kuulumattomista komponenteista poistuivat yllä kuvatun läpikäynnin ja muokkauksen myötä, mutta jäljelle jäivät jouduin piirtämään itse. Kävin läpi algoritmin järjestyksen 500 ensimmäistä merkkiä ja piirsin puuttuvat komponentit TakaoExMincho-fontin pohjalta.¹¹

5.3. Algoritmin tuloksen tarkastelua

Seuraavan sivun taulukossa on esitetty algoritmin luoman opiskelujärjestyksen 40 ensimmäistä merkkiä/komponenttia sekä vertailun vuoksi 40 käyttämäni korpuksen yleisintä sekä 40 Loachin ja Wangin (2016) määrittelemällä tavalla keskeisintä merkkiä. 300 ensimmäisen merkin lista on saatavilla liitteessä 1 ja kaikki jooyoo-kanjit järjestyksessä internetissä.¹² On helposti nähtävissä, että listat vastaavat varsin pitkälti toisiaan, vaikka algoritmin järjestyksessä on tietenkin seassa keskeisten merkkien komponentteja.

¹¹ Alkuperäinen fontti on saatavilla <https://launchpad.net/takao-fonts>. Ensimmäisten 500 merkin osalta jouduin piirtämään 17 merkkiä. Myöhemmin järjestyksessä tulevat 67 fontista puuttuvaa merkkiä ovat listassa edelleen ainoastaan Chise-tietokannan koodeilla.

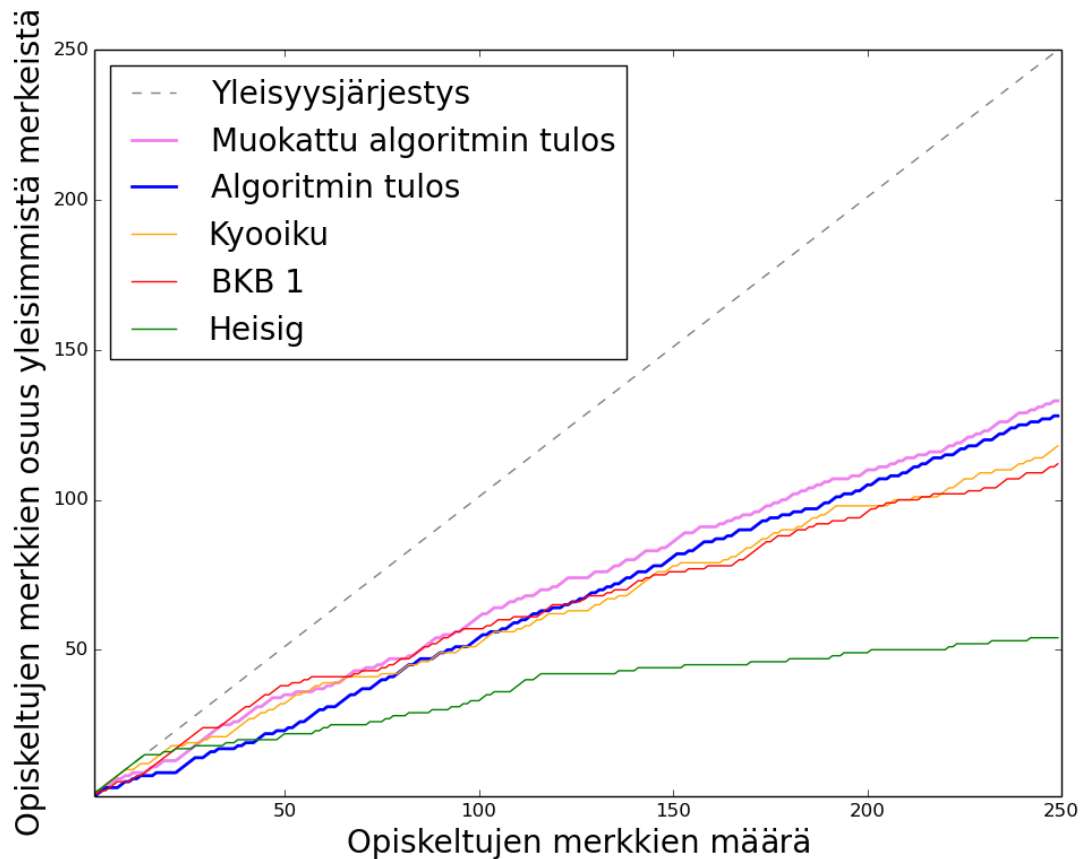
¹² https://ahaavisto.github.io/koko_kanjilista.html

	Algoritmin tulos BCCWJ-korpuksessa	Yleisimmät	Keskeisimmät
1	一 one	一	一
2	刀 sword	人	分
3	八 eight	大	人
4	分 part	出	行
5	人 person	分	間
6	丁 street	上	本
7	丁 to take small steps	中	時
8	彳 stop	見	的
9	行 going	生	大
10	日 day	行	思
11	門 gate	間	場
12	間 interval	自	上
13	木 tree	本	者
14	本 book	時	氣
15	寸 measurement	的	体
16	土 soil	合	実
17	寺 Buddhist temple	事	中
18	時 time	手	出
19	丶 dot	日	理
20	勺 wrapping enclosure	思	当
21	勺 ladle	前	部
22	丿 katakana no radical	方	動
23	白 white	場	会
24	的 bull's eye	年	対
25	大 large	言	要
26	心 heart	後	生
27	田 rice field	者	最
28	思 think	氣	二
29	勿 not	二	入
30	易 open	入	意
31	場 location	目	何
32	上 above	体	明
33	𠂇 variant of radical 125	実	持
34	者 someone	同	通
35	乂 ei käännöstä	立	手
36	乚 ei käännöstä	理	日
37	乂 ei käännöstä	当	多
38	气 spirit	部	家
39	氣 spirit	会	関
40	彳 radical number 9	動	取

Taulukko 1: Merkkilistojen alun vertailu. Harmaalla merkityt merkit esiintyvät jooyoo-kanjien joukossa ainoastaan komponentteina. Merkkilistan yhteydessä olevat englanninnokset on generoitu Kanjidic2-tietokannasta ottamalla aina kunkin merkin ensimmäinen käännös. Siten käännökset kuvaavat merkin merkityskenttää vain osittain.

Varsinaisten algoritmien alkuun nostamien merkkien lisäksi suuri osa niiden komponentteina esiteltävistäkin merkeistä on jooyoo-kanjeja (11 kappaletta), ja niistä seitsemän yleisimmän 500 kanjin joukossa (八, 門, 木, 土, 白, 心 ja 田). Heti alusta alkaen suurin osa opiskeltavista merkeistä on siis yleisiä ja siten hyödyllisiä kielenoppijalle. Opiskelujärjestyksen edetessä komponenttien osuus luonnollisesti vähenee, koska niistä monet ovat jo tuttuja aiemmista merkeistä.

Alla olevasta kuvasta näkyy, miten suuri määrä käyttämäni korpuksen yleisimmistä 250 merkistä on opiskeltu edetessä joko algoritmin mukaan, kyooiku-kanjien järjestyksessä, Basic Kanji Book -oppikirjan (Kanoh ym. 1990) järjestyksessä tai Heisigin järjestyksessä (Heisig 1977). Violetti viiva kuvaa algoritmilla luodun järjestyksen muokattua versiota, joka esitellään luvuissa 5.4–5.5. Tulos on varsin tasainen lukuun ottamatta Heisigin järjestystä, jossa merkkien yleisyyttä ei huomioida lainkaan, mutta algoritmi erottuu kuitenkin edukseen tällä tavalla yleisten merkkien opiskelun tehokkuutta mitatessa. Vielä selkeämpi ero tulee, jos huomioidaan vain algoritmin järjestystä seuraten opiskellut kanjit eikä niitä varten opiskeltuja jooyoo-merkkien ulkopuolisia komponentteja. 250 opiskellun merkin kohdalla algoritmin raakajärjestystä seuraten olisi opiskeltu jo miltei 200 yleisimmistä 250 merkistä. Tämä kuvaaja löytyy liitteestä 5.



Kuva 5: Kuvasta näkee, kuinka suuri osuus yleisimmästä 250 merkistä on opiskeltu missäkin vaiheessa muutamia eri opiskelujärjestyksiä seuraten.

Loachin ja Wangin algoritmilla on vastaavia rajoitteita kuten millä tahansa muullakin täysin laskennallisella optimointitavalla olisi. Oppimisen hinnan approksimoinnin vaikeuden mainitsin jo edellisessä luvussa, mutta itse algoritmin tulosta tarkastellessa käytetty approksimointi vaikuttaa tuottavan järjestyksen, joka noudattelee sille määriteltyjä tavoitteita yleisten merkkien opettamisesta havainnollisessa eli tässä tapauksessa komponenttiperusteisessa järjestyksessä. Tietojenkäsittelytieteilijät Loach ja Wang ovat jättäneet työn tähän vaiheeseen, eivätkä ole edenneet varsinaiseen lopputuloksen arviointiin lainkaan.

Algoritmin tulosta tarkastellessa käy ilmi, että se tuottaa toisinaan kummallisia approksimaatioita. Esimerkiksi ennen merkkiä 讓 'siirtää' opiskellaan 13 vedosta koostuva komponentti 襄 'auttaa', joka ei kuulu jooyoo-kanjeihin ja joka esiintyy komponenttina vain tässä yhdessä jooyoo-kanjissa. Vaikka muutin Chise-tietokannan komponenttijakoa poistaen juuri tämänkaltaisia turhia komponentteja, useissa tapauksissa niiden korvaaminen mielekkäillä alakomponenteilla olisi ollut hyvin työlästä tai jopa mahdotonta. Algoritmia jatkokehittäessä voisi olla syytä kehittää tapoja kiertää tämä turhien komponenttien ongelma muuttamalla laskukaavoja.

Isoimpana puutteena algoritmin tuloksessa kuitenkin on, että algoritmi ei ota lainkaan huomioon temaattisuutta järjestystä luodessaan: ihmisoppijalle hyödylliset sekä visuaaliset että semanttiset kategoriat jäävät täysin hyödyntämättä. Esimerkiksi ideografiset merkit 上 'yllä' ja 下 'alla' opiskellaan erillään toisistaan, samoin numeroita 1–10 vastaavat kanjimerkit. Varsinkin opiskelun alkuvaiheessa opiskelijoiden kanjitietämystä tukisi jonkinlaisten temaattisten kategorioiden hyödyntäminen kanjien opiskelujärjestyksessä. Tämän vuoksi algoritmin luomaa järjestystä on mielestäni välttämätöntä muokata hieman lisää, mikäli siitä halutaan kielen alkeisoppijalle käytännössä sopiva opiskelujärjestys.

5.4. Algoritmin tuloksen muokkaaminen käytettäväksi opiskelujärjestykseksi

Toteutin järjestyksen käsin muokkaamisen ensimmäisen noin 250 merkin osalta. Valitsin määrän vastaamaan Ippo-hankkeen lukion opetussuunnitelmassa määriteltyä lukion aikana opiskeltavien kanjien määrää (Ippo-hanke 2016). Toisaalta 250 merkkiä muodostaa kenelle tahansa opiskelijalle varsin hyvän pohjan, jonka jälkeen merkkien opiskelun jatkaminen onnistunee algoritmin luomassa raakajärjestyksessä.

Jaoin opiskeltavat merkit noin 15 merkin kokonaisuuksiksi, joita on yhteensä 16, tässäkin B3-kielen lukiokurssien määrä mielessäni: kullakin kurssilla voidaan opiskella kaksi kokonaisuutta. Varsinainen merkkimäärä jää tietenkin jonkin verran pienemmäksi, sillä osa algoritmin perusteella opiskeltavista merkeistä esiintyy jooyoo-kanjeissa pelkästään muiden merkkien komponentteina. Tässä esitettyä lukion aikana opiskeltavaa merkkimäärää olisi mahdollista siis nostaa, mikäli pelkkinä komponentteina opiskeltavia merkkejä ei lasketa mukaan merkkimäärään.

Otin järjestystä muokatessani huomioon kolme eri asiaa: temaattisten kategorioiden luomisen, yleisten yhdyssanojen molempien osien opiskelun samassa yhteydessä sekä sen, että kukin noin 15 merkin kokonaisuus olisi ainakin jossain määrin yhtenäinen osio, jonka merkeistä pystyy luomaan harjoituksia.

Opiskelujärjestystä olisi mahdollista muokata myös muokkaamalla merkkien laskennallisia keskeisyys-arvoja, eli lisäämällä etusijalle haluttujen merkkien kaavaan $\text{keskeisyys} = \text{frekvenssi/hinta}$ haluttu vakio z : $\text{keskeisyys} = \text{frekvenssi/hinta} + z$. En käyttänyt tätä lähestymistapaa tässä tutkimuksessa, vaan tein alustavan järjestyksen muokkaamisen käsin.

Temaattisten kategorioiden yhdistäminen algoritmin tulokseen

Edellä mainittujen hyötyjen ohella temaattisten kategorioiden esitleminen auttaa myös siinä, että opiskelluista merkeistä voidaan luoda mielekkäitä esimerkkilauseita ja -tekstejä eli esitellä merkkejä kontekstissaan. Kun merkin 日 *NICHI*, *hi* 'päivä; aurinko' kanssa yhdessä opiskellaan myös jonkin verran harvinaisempi (98. yleisin) merkki 月 *GATSU*; *tsuki* 'kuukausi; kuu', opiskelija oppii ilmaisemaan kokonaisia päiväyksiä japaniksi, esimerkiksi 8月6日, *hachigatsu muika*, '6. elokuuta'.

Jos merkkien opiskelujärjestyksestä haluaa hyötyä opintojen alusta alkaen, on syytä opettaa jonkin verran myös niitä merkkejä, joista on hyötyä muun oppimateriaalin lukemisessa ja muissa vastaantulevissa alkeistason teksteissä. Esimerkiksi merkit 学 'oppi' ja 校 'koulu', jotka muodostavat yhdessä sanan 学校 *gakkoo* 'koulu' kannattaa opettaa varhaisessa vaiheessa.

Selkein ero useimmissa olemassa olevissa merkkien opiskelujärjestyksissä verrattuna algoritmin raakatulokseen onkin nimenomaan temaattisuuden hyödyntäminen. Koska ihmisen muisti pyrkii luokittelemaan, on kanjien opiskelussa hyödyllistä käyttää luokittelua auttavia kategorioita muutenkin kuin sen osalta, että komponentit opiskellaan ennen niistä rakentuvia

yhdistelmämerkkejä.

Lähestyin kysymystä siitä, mitkä temaattiset kategoriat voisivat olla alkeisopiskelijalle hyödyllisiä tutustumalla joihinkin yleisesti käytettyihin opiskelujärjestyksiin. Niissä ilmenee tekijöidensä näkemys siitä, mitä alkeistasolla on hyödyllistä oppia. Kävin läpi listan kyooiku-kanjeista (MEXT 2017) ja kaksi yleisesti käytettyä JFL-oppikirjaa: *Basic Kanji Book 1* (Kanoh ym. 1990) JFL-kanjikirjan ja *Genki 1* (Banno, Ikeda ja Ohno 2011) JFL-oppikirjan,¹³ jossa kanjiopetus on integroitu mukaan oppikirjaan. Nämä kaikki, kuten useimmat muutkin kanjien opiskelujärjestykset, hyödyntävät temaattisuutta ainakin jossain määrin merkkien järjestysperusteena. Tunnetuista järjestyksistä lähinnä Heisig (1977) hyödyntää temaattisuutta ainoastaan minimaalisesti: aluksi esitellessään yhdessä esimerkiksi numerot ja muutamia muita pieniä kategorioita, kuten merkit 凹 'kovera' ja 凸 'kupera'. Wanikani-sovellus on modernimpi esimerkki temaattisuutta vain hyvin rajoitetusti hyödyntävistä oppimateriaaleista: siellä tiiviitä teemoja muodostavat ensimmäisen 86 kanjin joukossa ainoastaan 上 'ylös' ja 下 'alas', jopa perusluvut 1–10 esitellään kahdessa erillisessä vaiheessa.

kyooiku-kanjien ja *Genki*-kirjan osalta temaattisuus vähenee pidemmälle edetessä, mutta erityisesti *Basic Kanji Book* perustuu nimenomaan temaattisuuteen (Kano 1994). Näissä kaikissa kolmessa opiskelujärjestyksessä esiintyvät teemat ovat enimmäkseen hyvin konkreettisista asioista koostuvia semanttisia kategorioita, kuten eläimiä (鳥 'lintu', 魚 'kala'), luonnonmuotoja (川 'joki', 山 'vuori') tai sitten arkielämään kiinteästi liittyviä kategorioita, kuten viikonpäivien nimet (火 'tuli; tiistai', 水 'vesi; keskiviikko') ja numerot (一 'yksi', 五 'viisi'). Myös vastakohtaisuuteen liittyviä kategorioita hyödynnetään materiaaleissa: 右 'oikea' ja 左 'vasen' opetellaan näissä kaikissa järjestyksissä yhdessä, samoin 前 'ennen' ja 後 'jälkeen'.

Temaattisuus helpottaa – oletetun mutta vähän tutkitun helpomman omaksumisen ohella – kanjiharjoitusten luomista, sillä saman kategorian kanjeja saa yhdisteltyä helpommin mielekkäiksi lausekkeiksi, virkkeiksi, listoiksi tai esimerkiksi vastakohtapareiksi, jolloin opiskelija oppii samalla käyttökontekstia ja kanjisanojen suhdetta toisiinsa. Koska merkit esiintyvät todellisessa kielenkäytössä aina jossakin kontekstissa, on merkkien opiskelu ilman kontekstia varsin hyödytöntä.

Vaikka monet olemassa olevissa oppimateriaaleissa esiteltävistä merkeistä ovat hyvin yleisiä

¹³ Käytin itse oppikirjan sijaan oppikirjan virallisilla sivuilla olevaa listaa: <http://genki.japantimes.co.jp/self/genki-kanji-list-linked-to-wwkanji> (Luettu 5.2.2019)

kanjeja, niin temaattisuuden nimissä opitaan myös vähemmän hyödyllisiä merkkejä. kyooiku-järjestyksessä opitaan noin 120. merkin kohdalla vuorokauden aikoja, joista 昼 'päivä, keskipäivä' on paljon harvinaisempi (1091. BCCWJ-korpuksessa) kuin muut kategorian merkit 朝 'aamu' ja 夜 'ilta; yö'. Basic Kanji Bookin luvussa 5 puolestaan opiskellaan merkki 畑 'viljelty peltö'. Se ei kuitenkaan ole yleisimmän tuhannen merkin joukossa ja alkeisoppija törmää siihen vain kanjiharjoituksissa. Se on otettu mukaan vain, koska se koostuu loogisesti komponenteista 火 'tuli' ja 田 'peltö'. Sen mukaan ottamisen tarkoitus on opettaa opiskelijalle merkkien rakentumista komponenteista, mutta tässä tutkielmassa luodussa komponenttiperusteisessa opiskelujärjestyksessä komponentteja ei tarvitse tällä tavalla erikseen painottaa.

Koska tässä työssä luodun opiskelujärjestyksen päätarkoituksena on keskittyä nimenomaan opiskelemaan yleisiä merkkejä, valitsin vain pienen valikoiman teemoja mukaan. 254 ensimmäisen opiskeltavan merkin joukkoon otin mukaan kuusi kategoriaa:

- 1) numerot, kahdessa erässä 1–3 ja 4–10
- 2) suuntia: 左 vasen ja 右 oikea; 上 yllä, 下 alla ja 中 keskellä
- 3) määrän ilmauksia: 大 suuri ja 小 pieni; 多 paljon ja 少 vähän
- 4) viikonpäivät: 月 火 水 木 金 土 日
- 5) valikoiman opiskelun arkeen liittyvien sanojen merkkejä: 先生 'opettaja', 学生 'opiskelija', 学校 'koulu', 高校 'lukio', 質問 'kysymys', 問題 'ongelma', 語 'kieli', sekä muutoin arjessa hyvin tärkeä käsite 食べる 'syödä'
- 6) tärkeitä liikeverbejä: 行 'mennä', 来 'tulla', 帰 'palata', 出 '(mennä) ulos', 入 '(mennä) sisään'

Tämän sanaston ottaminen mukaan ensiksi opittaviin merkkeihin tuo myös konkretiaa opiskeltavaan sanastoon, sillä algoritmin tuloksessa painottuvat abstraktit käsitteet: kielen yleisimmät sanat ovat abstrakteja käsitteitä, joita käytetään kaikissa yhteyksissä. Näiden pienten temaattisten kategorioiden luomisen tai yhdistelyn lisäksi halusin ottaa huomioon myös useammasta kanjeista koostuvien, yleisten yhdyssanojen merkkien opettamisen yhdessä eräänlaisina parin merkin teemoina.

Yleisten yhdyssanojen huomiointi

Koska monet kanjit esiintyvät yleisesti nimenomaan tiettyjen yhdyssanojen osalta, on

oppijalle hyödyllistä opiskella yhtä aikaa kaikki yhdyssanan kanjit. Kävin läpi Balanced Corpus of Contemporary Written Japanese -korpuksen koko sanastosta tehdystä frekvenssilistasta (NINJAL 2015) yleisimpien 300 sanan joukossa esiintyneet kanjiyhdyssanat. Kyseisessä BCCWJ-korpuksen yleisessä frekvenssilistassa luetellaan lemmoja eikä huomioida, miten kukin sanan esiintymä on korpuksessa kirjoitettu, joten käytännössä monet yleisimmistä kanjiyhdyssanoista kirjoitetaan useimmiten hiraganalla. Esimerkiksi 何処 *doko* 'missä?' kirjoitetaan käytännössä aina nykykielessä hiraganalla どこ *doko*. Otin huomioon ne sanat, jotka yleensä kirjoitetaan nimenomaan kanjeilla, ja niiden osalta kaikki paitsi numeraaleista koostuvat yhdyssanat (kuten 二百 '200').

Raja 300 valikoitui sen mukaan, että aitoja kanjiyhdyssanoja alkaa frekvenssilistaa edetessä esiintyä jatkuvasti enemmän, ja 300 yleisimmän lemman joukossa esiintyi käsiteltäväksi sopiva määrä 38 yhdyssanaa. Kaikki ne koostuivat kahdesta kanjimerkistä. Mikäli algoritmin luomassa järjestyksessä esiintyi jonkin näistä yhdyssanoista osa, nostin myös yhdyssanan toisen osan opiskeltavaksi samassa kohdassa. Sain loogisesti mukaan noista yhdyssanoista 25, sillä jätin huomiotta sanat, joista kummankaan merkki ei esiinny algoritmin raakajärjestyksen ensimmäisten muutamien satojen merkkien aikana tai joihin alkeistason kielenoppija ei törmää missään, kuten 委員 *iin* 'komitean jäsen'.

Muut muokkaamisessa käytetyt strategiat

Päätin ottaa jokaiseen noin 15 merkin kokonaisuuteen mukaan yhden temaattisen kategorian, ja valita kategorian sillä perusteella, mitä merkkejä siinä vaiheessa algoritmin raakajärjestyksen perusteella opiskeltaisiin. Nostin tai laskin siis tiettyyn kokonaisuuteen haluttavien merkkien prioriteettia ja loin siten osittaista temaattisuutta opiskelujärjestykselle. Aiemmin listaamieni kuuden kategorian lisäksi osa kategorioita syntyi algoritmin valmiiksi lähes peräkkäin opiskeltaviksi nostamista merkeistä, jotka sattuivat muodostamaan jokseenkin yhtenäisen ryhmän (esimerkiksi teema *työelämä*). Kahdesta osiosta jäi tässä kohtaa teemaltaan ”sekalaisia”, sillä en halunnut muokata liikaa algoritmin luomaa järjestystä.

Otin huomioon myös opiskeltavien komponenttien sopimisen muihin kyseisen kokonaisuuden merkkeihin niiltä osin, kuin se oli helposti toteutettavissa. Esimerkiksi opiskeltaessa yhdyssanan 学校 'koulu' jälkimmäinen merkki 校 'koulu', opitaan sen komponentteina 交 'sekoittaa' ja sen alakomponenttina 父 'isä'. Tässä yhteydessä oli järkevää nostaa merkki 母 'äiti' mukaan samaan osioon: näin opiskelija oppii semanttisen kokonaisuuden 'vanhemmat'

molempien ydinkäsitteiden merkit samalla kertaa.

Siirsin opiskelujärjestystä muokatessani noin sata merkkiä eteenpäin järjestyksessä ja parikymmentä taaksepäin myöhemmin opiskeltavaksi. Suurinta osaa merkeistä siirrettiin vain hieman, sillä pyrin pitämään vahvasti kiinni algoritmin järjestyksen peruseriaatteista myös käsin järjestystä muokatessa.

5.5. Muokattu opiskelujärjestys

Opiskelujärjestykseni kokonaisuuksien teemat ovat järjestyksessä:

1. Ensimmäiset merkit; ajan ilmauksia
2. Viikonpäivät
3. Numeroita, sijainteja
4. Koulumaailma 1
5. Sekalaisia
6. Arkimaailma 1
7. Työelämä
8. Koulumaailma 2
9. Suunta ja määrä
10. Arkimaailma 2
11. Liikettä
12. Tietämistä, tuntemista
13. Värejä
14. Koulumaailma 3
15. Lisää ajanilmauksia
16. Sekalaisia 2

Osioiden keskinäinen järjestys perustuu pitkälti siihen, millaisia merkkejä opiskelujärjestyksen raakaversiossa oli missäkin kohdassa: samaan teemaan sopivia merkkejä on nostettu pidemmältä opiskelujärjestyksestä mukaan, samoin kuin hyödyllisiä yhdyssanoja muodostavia merkkejä. Temaattisuuden määrä vaihtelee eri osioiden välillä, sillä pyrin olemaan muuttamatta liikaa algoritmin luomaa, vahvasti frekvenssiin perustuvaa järjestystä.

Alla taulukossa ovat nähtävillä ensimmäiset neljä merkkikokonaisuutta. Koko muokattu 254 merkin opiskelujärjestys löytyy liitteestä 2.

Ensimmäiset kanjit	Viikonpäivät	Numeroita, sijainteja	Koulumaailma 1
一 one	月 month (nostettu)	四 four (nostettu)	𠂇 ei käännöstä (nostettu)
二 two (nostettu)	火 fire (nostettu)	五 five (nostettu)	子 child (nostettu)
三 three (nostettu)	水 water (nostettu)	六 six (nostettu)	学 study (nostettu)
刀 sword	金 gold (nostettu)	七 seven (nostettu)	生 life (nostettu)
八 eight	丁 street	九 nine (nostettu)	𠂇 kettle lid radical (nost.)
分 part	𠂇 to take small steps	十 ten (nostettu)	𠂇 upside-down box
人 person	𠂇 stop	勿 not	(nostettu)
日 day	行 going	易 open	高 tall (nostettu)
門 gate	大 large	場 location	儿 legs radical (nostettu)
間 interval	小 little (nostettu)	𠂇 ei käännöstä	𠂇 ei käännöstä (nost.)
木 tree	心 heart	口 mouth (nostettu)	先 before (nostettu)
本 book	田 rice field	合 fit (nostettu)	𠂇 ei käännöstä
寸 measurement	思 think	上 above	𠂇 ei käännöstä
土 soil		下 below (nostettu)	𠂇 ei käännöstä
寺 Buddhist temple		中 in (nostettu)	气 spirit
時 time			氣 spirit

Taulukko 2: Muokatun opiskelujärjestyksen ensimmäiset neljä kokonaisuutta eli ensimmäiset 59 merkkiä. Raakakäännökset Kanjidic2-tietokannasta. Harmaalla on merkitty ne komponentit, jotka eivät itsessään ole jooyoo-kanjeja, ja nostettu-merkinnällä ne, jotka on nostettu ensimmäisten 60 merkin joukkoon pidemmältä opiskelujärjestyksestä.

Näistä ensimmäisistä merkeistä 34 opiskellaan, koska ne ovat oleellisten merkkien komponentteja, mutta näistäkin suurin osa on suhteellisen yleisiä jooyoo-merkkejä. Mukana on myös useita monissa käsin luoduissa opiskelujärjestyksissä suosittuja piktografisia merkkejä, kuten 門 'portti', 木 'puu', 田 'riisipelto'. Ne toimivat hyvinä esimerkkeinä kanjimerkkien luonteesta sekä muodoltaan että käsitteeltään yksinkertaisesti omaksuttavista merkeistä: järjestys ei frekvenssille annetusta painoarvosta huolimatta käy liian abstraktiksi tai muutoin raskaaksi. Vaikka kanjien opiskelu aloitetaan monesti numeroiden 1–10 merkeistä, otin niistä mukaan vain ensimmäiset kolme. Nykyjapanissa käytetään yleisesti arabialaisia numeroita, joten numeroiden kanjien osaaminen ei ole välttämätöntä, jotta opiskelija pääsee tuottamaan ja lukemaan numeerisia ilmauksia. Niiden sijaan halusin heti aluksi näyttää kielessä yleisten substantiivien merkkejä.

Jo ensimmäisen kokonaisuuden merkeistä voidaan muodostaa paljon sanastoa, joka on yleistä sekä todellisessa elämässä että alkeisoppimateriaaleissa, kuten 分 *BUN* 'minuutti', 時 *JI* 'tunti', 時 *toki* 'aika', 日 *NICHI/hi* 'päivä', 一時間 *ichiJIKAN* 'yhden tunnin ajan' ja Japanin omankielinen nimi 日本 *nihon*. Seuraavan kokonaisuuden opiskeltuaan opiskelija tuntee kaikkien viikonpäivien nimet, ja kolmannen kokonaisuuden sanastolla (erityisesti 上 *ue* 'yllä'; 下 *shita* 'alla', 中 *naka* 'sisällä; keskellä') voidaan muodostaa joitakin tavallisia sijainteihin liittyviä ilmauksia. Osassa kolme nähdään myös ensimmäinen esimerkki yleisten yhdyssanojen huomioinnista. Merkki 場 *ba*, *JOO*, *CHOO* 'sijainti' esiintyy yleisimmin yhdyssanassa 場合 *ba'ai* 'tilanne', joten on mielekäästä nostaa korpuksen 16. yleisin merkki 合

a(u), *ai*, *GOO* 'tavata' (algoritmin raakajärjestyksessä 193.) samassa yhteydessä opiskeltavaksi.

Neljäs kokonaisuus on aiempia vielä selkeämmin temaattinen, sillä siellä opiskellaan kouluun ja opiskeluun liittyvistä sanoista kaikkein keskeisimpien merkit. Näin saadaan sanastoon mm. 学生 *GAKUSEI* 'opiskelija', 先生 *SENSEI* 'opettaja' ja 大学 *DAIGAKU* 'yliopisto'. Merkki 高 *taka(i)* *KOO* 'korkea' hyödyttää sekä adjektiivina *takai* 'korkea' että myöhemmin sanassa 高 (学)校 *KOO(GAK)KOO* 'lukio'.

6. Puoliautomaattisesti luotu oppimateriaali

Tukeakseni luomani kanjien opiskelujärjestyksen soveltamista käytäntöön olen luonut raakaversioita oppimateriaalista, joka noudattaa tätä järjestystä. Tämän työn laajuudessa en alkanut luomaan varsinaisia oppimista tukevia tehtäviä, vaan tein kanjeista internetiin listan esimerkikilauseineen sekä loin Anki-aikavälikertaussovellukseen virtuaalisen korttipakan, jonka avulla voi kerrata oppimiaan kanjeja. Listaa ja korttipakkaa voi tarkastella osoitteessa <https://ahaavisto.github.io/gradu.html>.

Aloitin oppimateriaalin luomisen generoimalla listan kanjeista esimerkikilauseineen ja tekemällä niin sanotun aikavälikertauksen korttipakan (luku 6.2). Ne ovat molemmat oikeastaan vain muistamisen apuvälineitä, ja oppija hyötyisi vielä niitakin enemmän tietenkin varsinaisesta oppimateriaalista lukuharjoituksine ja tehtävineen. Sellaisen luominen vaatisi kuitenkin niin paljon työtä, ettei se tämän opinnäytetyön puitteissa ole järkevää. Materiaalin eteenpäin kehittämistä pohditaan sekä tässä luvussa että Johtopäätöksiä-luvussa (luku 7).

6.1. Käännösten, ääntöasujen ja esimerkikilauseiden lisääminen puoliautomaattisesti

Generoituani algoritmilla kanjilistan ja muokattuani siitä käsin varsinaisen opiskelujärjestyksen, täytyi kanjilistan tueksi siihen lisätä vielä monia muita tietoja, jotta listasta tulisi opiskeluun sopiva apuväline. Lisäsin listaan KanjiDic2-tietokannasta kullekin merkille kaikki tietokannan tuntemat ääntöasut sekä kaikki englanninnokset, jotka tietokannassa valitettavasti eivät ole yhdistettynä niihin kuhunkin liittyviin ääntöasuihin vaan pelkästään kanjiin itseensä. Toteutin tämän jälleen python3-ohjelmointikielellä, kuten myös seuraavan vaiheen sanaston ja esimerkikilauseiden lisäämisen.

Saadakseni mukaan todellista japanin kielen sanastoa enkä ainoastaan yksittäisiä ääntöasuja ja käännöksiä ilman kontekstia, kirjoitin ohjelman yleisten kanjisanojen etsimiseen. Ohjelma luo ensin koko BCCWJ-korpuksen frekvenssilistasta erilliseen tiedostoon listan, jossa ovat kaikki ne sanat, joissa esiintyy kanjeja, ja jotka ovat frekvenssiltään kielen yleisimpien 5000 lemman joukossa. Tällaisia lemmoja listasta löytyi 4242 kappaletta. Sen jälkeen ohjelma liittää tästä listasta sanastoa kanjilistaan siten, että kuhunkin kanjiin liitettiin (enintään) yleisimmät viisi lemmaa. Esimerkiksi hyvin yleisen 日 'päivä; aurinko' merkin osalta liittämättä jäivät siten 26 seuraavaksi yleisintä lemmaa. Valitsin käyttämäni rajat muutaman kokeilun jälkeen sellaisiksi, että useimmille kanjeille (myös ainoastaan komponenttina esiintyville harvinaisille merkeille) löytyisi ainakin yksi esimerkkisana. Runsas määrä esimerkkisanoja tekee helpoksi valita niistä käsin alkeisoppijalle sopivimpia.

Seuraavaksi liitin kuhunkin lemmaan esimerkkilauseen, jossa lemma esiintyy. Tähän käytin Tatoeba-tietokannan¹⁴ niitä japaninkielisiä lauseita, johon oli liitetty suomenkielinen käännös. Tietokannasta löytyi 17 548 tällaista lausetta, joista kuitenkin osa on samojen lauseparien eri käännösversioita. Vapaaehtoisten luomassa tietokannassa esiintyy hyvin monenlaisia lauseita, josta on sekä hyötyä että haittaa: sopivimpien esimerkkilauseiden valinta vaatii paljon työtä, mutta toisaalta on mahdollista käyttää eri rekisterien esimerkkejä laajemmin kuin mitä monissa olemassaolevissa oppikirjoissa käytetään. Käännösten laatu on nähdäkseni pääosin erinomainen, vaikka myös heikkoja tai virheellisiä lauseita löytyy.

2) Esimerkkejä Tatoeba-tietokannan lauseiden variaatiosta

a) 心配しないで。 *shinpai naide.*

'Älä siitä huoli!' (edustava esimerkki puhekielisestä virkkeestä)

b) 大学時代の専攻は何でしたか？ *daigakujidai no senkoo wa nan deshita ka?*

'Mitä opiskelit pääaineenasi yliopistossa?'

(edustava esimerkki peruskohteliaan tyylin (*desu-masu*-tyylin) virkkeestä)

c) うちの犬がくしゃみするのは見たことあるけど、猫は一度もないな。

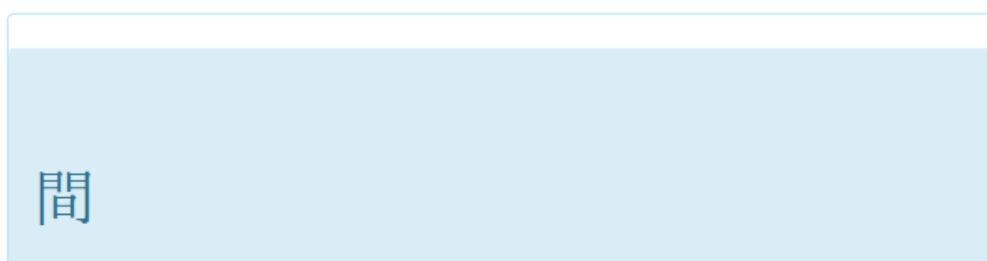
uchi no inu ga kushami suru no wa mita koto to aru kedo, neko wa ichido mo nai na

'Koiramme olen nähnyt aivastavan, mutta kissan en koskaan.'

(japaniksi oikeakielinen mutta merkitykseltään hieman omituinen esimerkki puhekielisestä lauseesta, joka on käännetty suomeen kirjakielelle)

¹⁴ <https://tatoeba.org/> (Luettu 5.1.2019). Tatoeba on Wiki-tyylinen tietokanta, johon kuka tahansa saa lisätä lauseita ja käännöksiä olemassaoleville lauseille. Käyttäjät voivat merkitä lauseita hyväksi tai huonoiksi, mutta mitään varsinaista asiantuntijamenettelyä tietokanta ei käytä.

Suurin ongelma Tatoeba-esimerkkilauseiden liittämisessä automaattisesti sanastoon oli, että käytin yksinkertaista hakufunktiota, joka etsi peräkkäisten kanjien esiintymiä ottamatta huomioon kielen rakennetta mitenkään. Esimerkiksi lemmalle 日 *hi* 'päivä; aurinko' yhdeksi esimerkkilausepariksi valikoitui 昨日食べすぎた *kinoo tabesugita* – *Söin liikaa eilen*, jossa merkki 日 esiintyy kahdella kanjilla kirjoitettavan sanan 昨日 *kinoo* 'eilen' osana, ja äännetään tässä tapauksessa poikkeuksellisella jukujikun-lukutavalla *noo*. Esimerkki on siis virheellinen, mutta jotta tietokoneohjelma saataisiin huomiomaan sanarajat ja merkkien ääntäminen kontekstissaan, tarvittaisiin ylimääräinen askel, *morfologinen jäsenitys* eli sanojen tunnistaminen. Koska japanissa ei käytetä sanavälejä, tämä on tietokoneelle vaativa tehtävä, ja käytännössä ohjelmaan pitäisi siksi yhdistää valmis morfologinen jäsenin, kuten esimerkiksi MeCab-sovellus¹⁵.



Komponentit: 門日

10. merkki

Kanjin merkityksiä englanniksi: interval, space

Kanjin lukutapoja: カン

ケン

あいだ

ま

あ

時間[じかん] (yleisyys: 134.)

Kymmenen vuotta on huomattava aika. 10年といえば相当な時間だ。

人間[にんげん] (yleisyys: 192.)

Ihmiset ovat kaikki tasa-arvoisia. 全ての人間は平等である。

間[あいだ] (yleisyys: 228.)

Isäni nukkui koko elokuvan ajan. 父は映画の間眠り通しだった。

Kuva 6: Ote internetiin luodusta kanjilistasta esimerkkilauseineen

¹⁵ <http://taku910.github.io/mecab/>. MeCab-jäsentimen algoritmeista sekä japanin morfologista jäsentimistä yleisesti ovat kirjoittaneet Kudo, Yamamoto ja Matsumoto (2004).

Tämän tutkielman puitteissa ratkaisin ongelman käsin: valitsin ohjelmani avulla esimerkkilauseetietokannasta kullekin lemmalle kolme esimerkkilauseetta ja kävin ne sitten itse läpi valiten niistä sopivimmalta vaikuttavan lauseen. Pyrin valitsemaan kullekin lemmalle lyhyen ja sanan merkitystä mahdollisimman hyvin kuvaavan esimerkkilauseen, ja mikäli yhtäkään sellaista ei ollut, poistin kaikki esimerkit. Samalla poistin listasta myös sanoja, jotka yleensä nykyjapanissa kirjoitetaan hiraganalla eikä kanjeilla, kuten 丁度, yleensä ちょうど *choodo* 'aivan niin', sekä kahdesti esiintyviä samoja ääntöasuja, joita tietokannassa esiintyi moneen sanaluokkaan laskettavissa olevien sanojen kohdalla.

Lauseiden valintaan niin algoritmisesti kuin käsinkin olisi jatkotutkimuksessa syytä kiinnittää tätä työtä enemmän huomiota. Pidän kuitenkin tärkeänä saada mukaan esimerkkejä kokonaisista lauseista, jotta kanjien opiskelijalla on mahdollisuus alusta asti nähdä merkkien todellista käyttökontekstia. Internetissä oleva kanjilistani on ensimmäinen suhteellisen laaja japani-suomi-esimerkkilauseiden kokoelma, joka on järjestetty kanjeihin perustuen.

6.2. Anki-aikavälikertauspakka

Tietoteknisiä apuvälineitä kanjien opiskeluun on tarjolla jatkuvasti enemmän, ja aihetta tutkitaan paljon. Tietotekniikan käyttämisen hyödyistä on monenlaista näyttöä. Esimerkiksi Librenjak, Vučković ja Dovedan (2012) raportoivat, että heidän JFL-opiskelijoille tekemässään kyselytutkimuksessa opiskelijat, jotka käyttivät tietotekniikkaa apuna kanjien opiskelussa, ilmoittivat hallitsevansa keskimäärin useampia kanjeja kuin muut opiskelijat. Morin, Omorin ja Saton (2016) tutkimuksessa todettiin japanin opiskelijoiden hyötyvän kanjien opiskelusta käännetyin luokkahuoneen (*flipped classroom*) metodilla, jossa kanjeja opiskeltiin itsenäisesti ennen oppituntia ja nimenomaan tietokoneavusteisin menetelmin. On kuitenkin tärkeä huomata, että Morin, Omorin ja Saton (2016) tutkimuksessa suurin hyöty oli opiskelijoiden kokemuksessa oppimisesta, ei oppimistuloksissa: mullistavia tuloksia tietotekniikan käyttöönotolla saavutetaan harvoin.

Ennen kaikkea tietotekniikan käyttö mahdollistaa oppimateriaalin eriyttämisen. Opiskelijat voivat esimerkiksi kerrata juuri niitä asioita, joissa kaipaavat tukea ja jättää vähemmälle huomiolle jo osaamansa asiat. Samalla mukaan saadaan pelillisyyttä, joka on iso oppimisen ja opettamisen trendi tällä hetkellä. Laadukkaiden oppimispelien luominen on työlästä, mutta pienessä määrin pelillisyyttä on helppo tuoda kanjiopetukseen virtuaalisten kertauskorttipakkojen avulla. Esimerkiksi (Mori, Omori ja Sato 2016) tutkimuksessa JFL-

opiskelijat pitivät enemmän korttipakan käytöstä kuin varsinaisista oppimispeleistä tai videoiden katsomisesta.

Virtuaaliset kertauskortit perustuvat jo 1940-luvulla hyödynnettyyn tekniikkaan: kortin toiselle puolelle kirjoitetaan tuttu asia ja toisella puolella opittava, esimerkiksi suomenkielinen sana ja sen japaninkielinen käännös. Oppija yrittää korttia katsoessaan muistaa, mitä toisella puolella lukee, kääntää kortin ja asettaa sen tiettyyn pinoon sen mukaan, muistiko kysytyn asian vai ei. Ne kortit, joita ei muistettu kerrataan heti ja muistettuja asioita kerrataan harvenevilla aikaväleillä. (Godwin-Jones 2010, 7). Nykyaikana korttipakat ovat yleensä virtuaalisia ja niitä selataan tietokoneella, mutta periaate on sama. Virtuaalisuus mahdollistaa sellaista, mihin pahvikortit eivät taivu: kertausaikavälien joustavan säädön, tilastojen näyttämisen käyttäjälle sekä korttien automaattisen luomisen, kunhan pohjadata on olemassa.

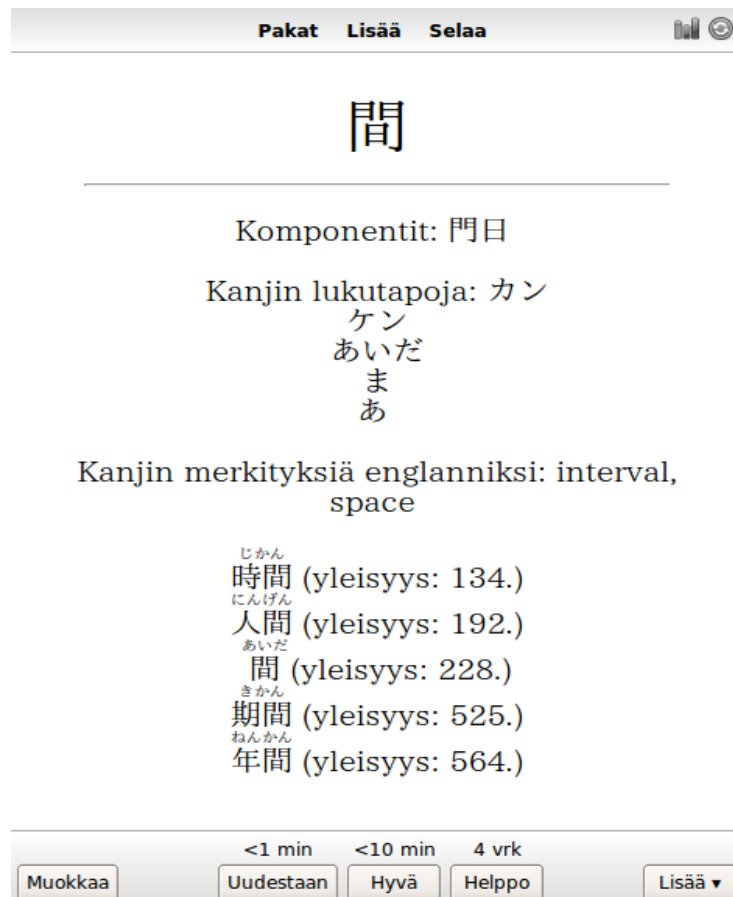
Toteutin luomaani opiskelujärjestykseen perustuvan aikavälikertaus-korttipakan Anki-ohjelmaan¹⁶, sillä se on sekä laajasti käytetty että minulle entuudestaan tuttu. Ankin voi ladata ilmaiseksi omalle Windows-, Mac- tai Linux-koneelle sekä Android-mobiililaitteelle. Applen mobiililaitteille ohjelmasta on saatavilla maksullinen versio, mutta sovelluksen ilmaista internetselainversiota¹⁷ voi käyttää myös niillä.

Aikavälikertauspakkoja voi luoda Ankiin sekä käsin kortti kerrallaan että tuomalla ohjelmaan määrämuotoisen tekstitiedoston. Tämän tutkimuksen dataa käyttävän korttipakan loin luomalla tarvittavan tekstitiedoston python3-ohjelmointikielellä ja tuomalla sen sitten kokonaisena ohjelmaan. Liitin pakkaan samat tiedot kuin kanjilistaan: merkkien komponentit, merkin ääntöasut, listan mahdollisia käännöksiä englantiin ja enintään viisi yleistä sanaa. Esimerkkilauseita en kuitenkaan lisännyt, jotta yhdestä kortista ei tulisi liian pitkää.

Aikavälikertaus sopii yhdeksi oppimisen apuvälineeksi, mutta ei toimi yksinään. Sen avulla on vaikea tuoda kontekstia oppimiseen, ja kyseessä onkin loppujen lopuksi ainoastaan hieman optimoitu muoto perinteisille toistomenetelmille. Useimmissa ohjelmissa ei myöskään ole mahdollisuutta piirtää itse kirjoitusmerkkejä, joten kirjoitustaidon harjoittelu pitää tehdä muualla. Suuri osa kanjien opiskelusta on kuitenkin yksinkertaisesti merkin muodon, merkityksen ja äännearvon muistamista, ja siihen aikavälikertaus soveltuu.

¹⁶ <https://apps.ankiweb.net/>

¹⁷ <https://ankiweb.net/>



Kuva 7: Näkymä Anki-ohjelmasta, kun kortti “間” on ”käännetty” eli kanjin lisäksi näkyvissä ovat itse merkin ohella muut merkkiin liittyvät tiedot sekä lista yleisimmistä sanoista.

Pakkaa voisi kehittää vielä monin tavoin. Kuten kanjilistankin osalta, ensimmäiseksi olisi syytä tarkistaa sen tiedot sekä lisätä suomennoksia. Esimerkiksi kunkin merkin vetomäärän ja radikaalin lisääminen onnistuisi Kanjidic2-tietokannasta. Esimerkkisanastosta ja -lauseista voisi luoda omat korttipakkansa. Tämän työn puitteissa en kuitenkaan lähtenyt laajentamaan Anki-pakkaa tämän enempää. Jo nyt pakkaa voisi käyttää muun opiskelun tukena merkkien ääntöasujen ja merkitysten muistamiseen.

7. Yhteenveto

Tässä tutkimuksessa kuvailin Loachin ja Wangin (2016) luoman algoritmin perusteella japanin kielen kirjoitusmerkkien opiskelujärjestyksen luomisen laskennallisesti ja tutkin luotua järjestystä. Kehitin järjestystä eteenpäin kohti todellisuudessa hyödynnettäväksi sopivaa järjestystä muokkaamalla sitä käsin mielekkäämpään suuntaan, sillä algoritmin tulokseen tutustuessi oli selvää, että teoriassa erinomainen järjestys ei ole välttämättä käytännössä erinomainen.

Algoritmin luoman raakaversiota tarkastellessa selkein puute oli temaattisuuden puuttuminen kokonaan: esimerkiksi numeroiden 1–10 kanjimerkit opiskellaan 440 ensimmäisen merkin opiskelun mittaan erillään toisistaan. Päätin siksi luoda käsin optimoidun järjestyksen ensimmäiselle noin 250 merkille. Järjestystä luodessani pidin huolta temaattisuudesta myös siinä mielessä, että tutkin algoritmin alussa opiskeltavaksi nostamien sanojen esiintymistä yhdyssanoissa ja yhdistin yleisiä yhdyssanoja muodostavia kanjeja opiskeltavaksi samaan aikaan.

Tämä muokattu opiskelujärjestys mahdollistaa monipuolisten opetus- ja opiskelustrategioiden käytön. Etymologiaa, vetojärjestystä ja kanjien perusrakennetta on helppo käsitellä komponenttien kautta, ja komponenttien tunteminen tukee myös muistisääntöjen keksimisessä, mikäli opiskelija haluaa hyödyntää sellaisia. Vastaavasti merkkejä pystyy myös ryhmittelemään niin semanttisten kuin foneettistenkin komponenttien perusteella. Merkkejä on helppo tuoda esille autenttisisessa kontekstissa, sillä opiskellut merkit ovat yleisiä ja esiintyvät kaikenlaisissa teksteissä. Näin siis Paxtonin ja Svetanantin (2013, 95–97) mainitsemista kolmesta perusopiskelustrategiasta (toisto, muisti, konteksti) kaikkia pystyy hyödyntämään tehokkaasti. Myös esimerkiksi Gamagen (2003) kyselytutkimuksen tarkemmin jaotelluista 20 strategiasta pystyy luontevasti käyttämään lähes kaikkia.

Algoritmin luoman järjestyksen pohjalta muokattu järjestys ottaa edelleen huomioon periaatteet mahdollisimman yleisten eli hyödyllisten merkkien opiskelusta. Sekä algoritmin raakajärjestys että sen pohjalta muokattu opiskelujärjestys opettavat 250 ensimmäisen merkin aikana suuremman osan korpuksen yleisimmistä merkeistä kuin muut olemassa olevat opiskelujärjestykset, joihin sitä verrattiin (ks. kuva 5). Samanaikaisesti periaate helposti omaksuttavasta järjestyksestä tulee huomioitua järjestyksen hierarkkisuu den vuoksi, ja temaattisuus antaa opiskelijalle lisätyökaluja merkkien ja niiden muodostaman sanaston kokonaisuuden hahmottamiseen.

Vaikka opiskelujärjestys on uusi, sekä raakajärjestys että muokattu järjestys opettavat suurelta osin samoja merkkejä kuin ensikielen opiskelijoiden käyttämä kyooiku-järjestys tai useimmissa yleisissä JFL-oppikirjoissa käytetyt järjestykset. Tämä kertoo siitä, että perusteet merkkien järjestämiseen ovat varsin yhtenäisiä. Suurimpana erona on, että algoritmisesti luotu järjestys huomioi jatkuvasti merkkien komponentit, joten esimerkiksi Basic Kanji Bookin (Kano h ym. 1990) ja useiden muiden oppikirjojen käyttämää strategiaa joidenkin merkkien

esittelystä vain esimerkkinä komponenteista ei tarvita, vaan opiskelujärjestyksessä voidaan jatkuvasti esitellä hyödyllisiä merkkejä.

Vaikka Loachin ja Wangin (2016) parametrisoivat ihmisen kognitiota ja kielen oppimista ilman perusteluja, tässä tutkielmassa tehty tarkempi tarkastelu osoittaa heidän parametriensa ja algoritminsa muodostavan mielekkään pohjan japanin kielen opiskelujärjestyksen kehittämiseksi. Kanjimerkkien opiskelujärjestyksen luonti vaatii yhä paljon ihmistyötä, mutta sekä tässä työssä hyödynnetyt korpuksat ja tietokannat että itse algoritmi ovat osoittautuneet varteenotettaviksi työkaluiksi.

Algoritminen optimointi ja muut laskennalliset menetelmät voivat hyödyttää kanjien opiskelua ja oppimateriaalin luomista monin tavoin. Alaa ei ole vielä tutkittu paljon, ja tämäkin tutkielma herättää monia jatkokysymyksiä. Tärkeä jatkotutkimuskohde olisi tutkia käytännössä nyt luodun opiskelujärjestyksen vaikutusta oppimiseen. Jotta järjestyksen toimivuudesta saisi luotettavaa tietoa, pitäisi tämän kaltainen tutkimus toteuttaa huolellisesti, kahta alkutilanteeltaan mahdollisimman identtistä ryhmää vertaillen. Koska työn tarkoituksena on ollut ulottaa optimoitu opiskelujärjestys koko jooyoo-kanjilistaan, tutkimusta olisi hyvä jatkaa muutaman kuukauden sijaan ainakin vuoden tai kahden ajan.

Olisi myös kiinnostavaa vertailla, kuinka paljon opiskelujärjestys muuttuisi, mikäli se optimoitaisiin yleisimpien kanjisanojen (jotka ovat usein yhdyssanoja) eikä pelkkien merkkien yleisyyden mukaan. Esimerkiksi Loach ja Wang (2016) toteuttivat tutkimuksessaan vastaavan vertailun, mutta japanin kielestä ei löytynyt sopivaa koneluettavaa frekvenssilistaa, josta olisi saanut nimenomaan kanjisanojen määrän. Monet frekvenssilistat, kuten koko BCCWJ-korpuksesta luotu lista (NINJAL 2015) listaavat lemmat ottamatta huomioon sitä, että sama lemma voidaan monesti kirjoittaa japanin kielessä useammalla tavalla (esim. 事 ja こと, 'asia'). Siten listasta ei saa tietoa kanjien yleisyydestä.

Temaattisuuden huomiointia voisi kehittää laskennallisilla menetelmin. Ihmisen valitsemien merkkien nostaminen opiskelujärjestyksessä aiemmin opiskeltaviksi onnistuu algoritmisesti jo nyt muokkaamalla kanjien keskeisyys-arvoja käsin, mutta teemoja voisi etsiä myös sanojen merkityskenttiä ja suhteita tutkivilla laskennallisilla menetelmillä, kuten Word2Vec-algoritmillä (Mikolov, Chen, Corrado ja Dean 2013). Laskennallisista menetelmistä olisi hyötyä erityisesti silloin, jos koko 2136 merkin opiskelujärjestystä haluttaisiin muokata.

Nyt luotua oppimateriaalia voi kehittää monin tavoin. Tällä hetkellä sinne on liitetty kaikki

tietokannan tuntemat ääntöasut ja englanninnokset, mutta niistä olisi syytä valita ainoastaan yleiset tai muutoin havainnolliset. Käsien tarkistamisen lisäksi laskennallisia menetelmiä voisi hyödyntää muun muassa esimerkkilauseiden valinnassa vielä paljon. Ensimmäiseksi olisi järkevää korjata se, että tietyille lemmalle esimerkkilauseetta lisätessä ohjelma ei ota huomioon merkin kontekstia ja ääntämystä kontekstissa, eli käytännössä ohjelmaan olisi syytä yhdistää morfologinen jäsenin (ks. tarkemmin luku 6.1).

Tässä tutkielmassa kartoitettiin mahdollisuuksia laskennallisten algoritmien, valmiiden kanji- ja sanakirjatietokantojen sekä korpusten hyödyntämistä kanjien opiskelujärjestyksen luonnissa. Vaikka tulokset ovat alustavia, ja ihmistä tarvitaan yhä tekemään lopulliset valinnat opiskeltavista merkeistä sekä tarkistamaan automaattisesti generoidut tiedot, kaikki käytetyt menetelmät osoittautuivat lupaaviksi. Tietotekniikan ja saatavilla olevien aineistojen kehittyessä mahdollisuudet laajenevat entisestään.

8. Lähteet

- Araki-Masonen, Kumi. 2006. *Hiragana ja katakana: opas japanilaisen tavukirjoituksen maailmaan*. Helsinki: Finn Lectura.
- Banno, Eri, Yoko Ikeda, ja Yutaka Ohno. 2011. *GENKI I: An integrated course in elementary Japanese*. 2. painos. Tokio: Japan Times/Tsai Fong Books.
- Bruneau, Päivi. 2009. *Suuri kanjikirja*. Jyväskylä: Atena.
- Carroll, Tessa. 2001. *Language planning and language change in Japan*. Richmond, Surrey: Psychology Press.
- Chikamatsu, Nobuko. 2005. Japanese kanji memory and retrieval: An experiment on the tip-of-the-pen (TOP) phenomenon. Teoksessa *Second Language Writing Systems*, toimittanut Vivian Cook ja Benedetta Bassetti, 71–96.
- CHISE (CHAracter Information Service Environment) project. 2018. ”IDS-UCS-Basic.txt”. Luettu 27.8.2018. http://git.chise.org/gitweb/?p=chise/ids.git;a=blob_plain;f=IDS-UCS-Basic.txt.
- Electronic Dictionary Research and Development Group. ei pvm. KanjiDic2. Luettu 31. elokuuta 2018. <http://www.edrdg.org/kanjidic/kanjidic2.xml.gz>.
- Everson, Michael E. 2011. Best practices in teaching logographic and Non-Roman writing systems to L2 learners. *Annual Review of Applied Linguistics* 31: 249–274.
- FSI’s experience with language learning. Ei pvm. Luettu 2. tammikuuta 2019. <https://www.state.gov/m/fsi/sls/c78549.htm>.
- Gamage, G Haththotuwa. 2003. Perceptions of kanji learning strategies: Do they differ among Chinese character and alphabetic background learners? *Australian Review of Applied Linguistics* 26 (2): 17–31.
- Godwin-Jones, Robert. 2010. Emerging technologies from memory palaces to spacing algorithms: approaches to secondlanguage vocabulary learning. *Language, Learning & Technology* 14 (2): 4–11.
- Halpern, Jack. 1990. *New Japanese-English character dictionary*. 1st edition. Tokio: Kenkyusha.
- Heisig, James W. 1977. *Remembering the kanji I: A complete course on how not to forget the meaning and writing of Japanese characters vol. 1*. Tokio: Japan Publications Trading Co.
- Henshall, Kenneth G. 1989. *A Guide to remembering Japanese characters*. Tokio: Tuttle Publishing.
- Ilmonet. 2018. Luettu 4. syyskuuta 2018. <https://ilmonet.fi>.
- Ippo-hanke. 2016. B3-Japanin OPS. <https://ippohanke.wordpress.com/2016/05/02/b3-japanin-ops/>.
- Iwasaki, Shoichi. 2013. *Japanese: Revised edition*. 2. painos. London Oriental and African Language Library 21. Amsterdam: John Benjamins Publishing Company.
- Japan Foundation. 2017. Survey report on Japanese-language education abroad 2015.
- Kanoh, Chieko, Hiroko Takenaka, Eriko Ishii, ja Yuri Shimizu. 1990. *Basic kanji book, Vol. 1*. Tokio: Bonjinsha.
- Kanoo, Chieko. 1994. 漢字教育のためのシラバス案 (Kanji kyooiku no tame no shirabasu an) [Opetusohjelma kanjiopetukseen]. *Journal of Japanese Language Teaching* 9: 41–50.
- Kardy, Glenn. 2007. *Kana de manga: Helppo ja hauska tapa oppia ”japanilaiset aakkoset”*. Helsinki: Sangatsu manga.

- Kardy, Glenn, ja Chihiro Hattori. 2008. *Kanji de manga: helppo ja hauska tapa oppia japania. Osa 1*. Helsinki: Sangatsu Manga.
- Karppinen, Takako. 1999. *Japanin kielen alkeet*. Helsinki: Finn Lectura.
- Kess, Joseph F., ja Tadao Miyamoto. 2000. *Japanese mental lexicon: Psycholinguistic studies of kana and kanji processing*. Amsterdam: John Benjamins Publishing Company.
- Kielitiede:grafeemi – Tieteen termipankki. ei pvm. Luettu 26. tammikuuta 2019.
<http://tieteentermipankki.fi/wiki/Kielitiede:grafeemi>.
- Kudo, Taku, Kaoru Yamamoto, ja Yuji Matsumoto. 2004. Applying conditional random fields to Japanese morphological analysis. Teoksessa *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 230–237.
- Librenjak, Sara, Kristina Vučković, ja Zdravko Dovedan. 2012. Multimedia assisted learning of Japanese kanji characters. Teoksessa *Proceedings of the 35th International Convention*, 1470–1475.
- Loach, James C., ja Jinzhao Wang. 2016. Optimizing the learning order of Chinese characters using a novel topological sort algorithm. Toimittanut Xuchu Weng. *PLOS ONE* 11 (10).
- Lu, Mei-Yan, Krus, David J., James M. Webb, ja Laura S. Fox. 1999. Using order analytic hierarchies of mnemonics to facilitate learning of Chinese and Japanese kanji characters. *The Journal of Experimental Education* 67 (4): 293–311.
- Matsumoto, Kazumi. 2013. Kanji recognition by second language learners: Exploring effects of first language writing systems and second language exposure. *The Modern Language Journal* 97 (1): 161–177.
- MEXT. 2017. 小学校学習指導要領 (Shoogakkoo gakushuu shidoo yoorei) [Alakoulun opetussuunnitelman perusteet]. Monbukagakusho.
- Mikolov, Tomas, Kai Chen, Greg Corrado, ja Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. Teoksessa *Proceedings of the International Conference on Learning Representations*.
- Mori, Yoshiko. 2014. Review of recent research on kanji processing, learning, and instruction. *Japanese Language and Literature* 48 (2): 403–430.
- Mori, Yoshiko, Motoko Omori, ja Kumi Sato. 2016. The impact of flipped online kanji instruction on written vocabulary learning for introductory and intermediate Japanese language students. *Foreign Language Annals* 49 (4): 729–749.
- Nichiguchi, Koichi. 2001. *Minna no nihongo: Kanji I*. Tokio: 3A Corporation.
- NINJAL (kokuritsu kokugo kenkyuujo). 2011. NDC Genre-specific Kanji Frequency List.
https://pj.ninjal.ac.jp/corpus_center/bccwj/data-files/frequency-list/NDC-Kanji-freqlist.zip.
- . 2015. BCCWJ Word List Balanced Corpus of Contemporary Written Japanese.
http://pj.ninjal.ac.jp/corpus_center/bccwj/en/freq-list.html.
- Nozaki, Hironari, ja Yasutaka Shimizu. 2000. An analysis of kanji frequency in newspaper articles, and the development of a kanji list for teaching newspaper Japanese. *Japan Journal of Educational Technology* 24 (2): 121–132.
- Nozaki, Hironari, Shoichi Yokoyama, Yukuo Isomoto, ja Junko Yoneda. 1996. A study of character frequency : From the view point of Japanese language education. *Japan Journal of Educational Technology* 20 (3): 141–149.
- Okura, Junichiro. 2011. *Elävää japania. 1*. Helsinki: Finn Lectura.
- Opetushallitus. 2015. Lukion opetussuunnitelman perusteet 2015. Opetushallitus.
http://www.opi.fi/download/172124_lukion_opetussuunnitelman_perusteet_2015.pdf.
- Paxton, Simon, ja Chavalin Svetanant. 2013. Tackling the kanji hurdle: Investigation of kanji

- learning in non-kanji background learners. *International Journal of Research Studies in Language Learning* 3 (3).
- Richmond, Stephen. 2005. A re-evaluation of kanji textbooks for learners of Japanese as a second language. *Journal of the Faculty of Economics, KGU* 15: 43–71.
- Romera Martinez, Tao. 2016. IiKanji: Optimized kanji reading learning method. 漢字学習研究会誌 (Kanji gakushuu kenkyuukai shi) 5: 129.
- Rose, Heath. 2013. L2 learners' attitudes toward, and use of, mnemonic strategies when learning Japanese kanji: L2 learners' attitudes toward, and use of, mnemonic strategies. *The Modern Language Journal* 97 (4): 981–992.
- Seeley, Christopher. 1991. *A history of writing in Japan*. Leiden: Brill.
- Shen, Helen H., ja Chuanren Ke. 2007. Radical awareness and word acquisition among nonnative learners of Chinese. *The Modern Language Journal* 91 (1): 97–111.
- Shimizu, Hideko, ja Kathy E. Green. 2002. Japanese language educators' strategies for and attitudes toward teaching kanji. *The Modern Language Journal* 86 (2): 227–241.
- Summary: Balanced Corpus of Contemporary Written Japanese. 2009. Luettu 30.9.2018. http://pj.ninjal.ac.jp/corpus_center/bccwj/en/.
- Tamaoka, Katsuo. 2014. The Japanese writing system and lexical understanding. *Japanese Language and Literature* 48 (2): 431–471.
- Tamaoka, Katsuo, ja Sachiko Kiyama. 2013. The effects of visual complexity for Japanese kanji processing with high and low frequencies. *Reading and Writing* 26 (2): 205–223.
- Tatoeba: Collection of sentences and translations. Viitattu 5. tammikuuta 2019. <https://tatoeba.org/eng>.
- Taylor, Insup, ja M. Martin Taylor. 2014. *Writing and literacy in Chinese, Korean and Japanese: Revised edition*. Amsterdam: John Benjamins Publishing Company.
- Tofugu, nimimerkki Koichi. 2018. Sähköpostiviesti, 22. maaliskuuta 2018.
- Toyoda, Etsuko. 2000. English-speaking learners' use of component information in processing unfamiliar kanji. *Australian Review of Applied Linguistics* 23 (1): 1–14.
- Tuomi, Tapio J. 1998. *Japanilais-suomalainen kanji-sanakirja*. Helsinki: Japanilaisen kulttuurin ystävät.
- Vorobev, Victor, ja Galina Vorobeva. 2012. An analysis of the efficiency of existing kanji indexes and development of a coding-based index. *Acta Linguistica Asiatica*: 2(3), 27–59.
- Vorobeva, Galina. 2011. 構造分析とコード化に基づく漢字字体情報処理システムの開発 (Koozoo bunseki to koodoka ni motozugu kanjijikei joohoo shori shisutemu no kaihatu) [Informaationprosessointijärjestelmän kehittäminen kanjien muodoille perustuen rakenteen analyysiin ja koodaukseen]. 日本語教育 (Nihongokyoouiku): 145, 16–30.
- . 2014. 構造分解とコード化を利用した計量的分析に基づく漢字学習の体系化と効率化 (koozoo bunseki to koodoka wo riyuu shita keisanteki bunseki ni motozuku kanji gakushuu no taikenka to sotsuritsuka) [Kanjien rakenteen laskennalliseen analyysiin ja koodaukseen perustuva kanjien järjestäminen]. Tohtorinväitös. <http://www.grips.ac.jp/cms/wp-content/uploads/2014/09/20140909Vorobyova-diss3beznachipustyhstr.pdf>.
- Vorobeva, Galina, ja Victor Vorobev. 2017. 非漢字系日本語学習者の漢字学習における阻害要因とその対処法: 体系的な漢字学習の支援を目指して (Hikanjikei nihongo gakushuusha no kanji gakushuu ni okeru sogai yooiin to sono taishohoo: taiketeikina kanjigakushuu no shien wo mezashite) [Ei-kanjitaustan opiskelijoiden kanjien opiskelun esteet ja niiden karsiminen]. *NINJAL Research Papers* 12: 163–179.

- Wanikani. ei pvm. Luettu 8. tammikuuta 2018. <https://wanikani.com/>.
- Yamashita, Hiroko, ja Yukiko Maru. 2000. Compositional features of kanji for effective instruction. *The Journal of the Association of Teachers of Japanese* 34 (2): 159–178.
- Yan, Xiaoyong, Ying Fan, Zengru Di, Shlomo Havlin, ja Jinshan Wu. 2013. Efficient learning strategy of Chinese characters based on network approach. Toimittanut Matjaz Perc. *PLoS ONE* 8 (8).
- Yanzu-hanke. 2012. 5.5.6 Vieraat kielet, Aasian ja Afrikan kielet, B3-oppimäärä (Kiina). http://yanzu.fi/wp-content/uploads/2012/01/Kiinan-B3lops_kurssikuvaukset.pdf.
- 常用漢字表 (jooyoo kanji hyoo) [Lista yleisessä käytössä olevista kanjeista]. 2010. Agency of Cultural Affairs, Government of Japan. http://www.bunka.go.jp/kokugo_nihongo/sisaku/joho/joho/kijun/naikaku/pdf/joyokanjihyo_20101130.pdf.

Liitteet

Liite 1: Algoritmin tuottama opiskelujärjestys

Ensimmäiset 300 merkkiä komponentteineen algoritmin määrittelemässä opiskelujärjestyksessä. Koko 2136 jooyoo-kanjin ja 503 muun komponentin lista on nähtävillä internetissä osoitteessa http://ahaavisto.github.io/koko_kanjilista.html.

一 一	丿 丿	生 生	戊 戊
刀 刀	乚 乚	又 又	成 戊 丁
八 八	气 乚 一 丿	耳 耳	牛 牛
分 八 刀	氣 气 乂	取 耳 又	物 牛 勿
人 人	イ イ	最 日 取	方 方
丁 丁	体 イ 本	入 入	疋 疋
丁 一 丁	夫 夫	音 音	定 ㄣ 疋
彳 彳	ㄣ ㄣ	意 音 心	矢 丿 天
行 彳 丁	実 ㄣ 夫	丁 丁	知 矢 口
日 日	中 中	可 丁 口	自 自
門 門	出 出	何 イ 可	𠂇 𠂇
間 門 日	里 田 土	月 月	性 𠂇 生
木 木	王 王	明 日 月	玉 、 王
本 木 一	理 王 里	𠂇 𠂇	口 口
寸 寸	𠂇 𠂇	持 𠂇 寺	国 口 玉
土 土	𠂇 𠂇	用 用	𠂇 𠂇
寺 土 寸	当 𠂇 𠂇	マ マ	𠂇 𠂇
時 日 寺	𠂇 𠂇	甬 マ 用	𠂇 𠂇
、 、	口 口	𠂇 𠂇	𠂇 𠂇
勺 勺	立 立	通 𠂇 甬	𠂇 𠂇
勺 勺 、	音 立 口	手 手	作 イ 𠂇
丿 丿	部 音 𠂇	夕 夕	也 也
白 丿 日	力 力	多 夕 夕	地 土 也
的 白 勺	重 重	豕 豕	舌 舌
大 大	動 重 力	家 ㄣ 豕	言 言
心 心	ム ム	天 一 大	話 言 舌
田 田	二 二	𠂇 𠂇	ラ ラ
思 田 心	云 二 ム	关 𠂇 天	今 ハ ラ
勿 勿	ハ ハ	関 門 关	弋 弋
易 勿 日 一	会 ハ 云	見 見	代 イ 弋
場 土 易	文 文	現 王 見	七 七
上 上	対 文 寸	全 ハ 王	化 イ 七
𠂇 𠂇	女 女	𠂇 𠂇	禾 禾
者 𠂇 日	西 西	考 𠂇 𠂇	私 禾 ム
× ×	要 西 女	丁 丁	下 下
			ト ト

外夕卜
年年
儿儿
牛牛
先牛儿
子子
斤斤
戸戸
所戸斤
目目
小小
少小ノ
久久
亦亦
変亦久
学子
事主、王
表去
法去
聿聿
書聿日
三三
么么
戈戈
幾戈么人
機木幾
問門口
十十
未未
味口未
亦立亦
亲亲斤
新亲斤
欠欠

次欠
升升
开一升
開門开
士士
吉士口
糸糸
結糸吉
名夕口
皮皮
彼夕皮
合ハ一口
咸口戌一
感咸心
無無
無無
示示
社ネ土
近近斤
止止
正一止
内内
弓弓
引弓
リリ
前月
吏吏
使イ吏
特牛寺
青丰月
情卜青
車車
連車
果田木
其其

其其八
期其月
売士儿
続糸売
後イ久么
ネネ
初ネ刀
頁頁
是日正
題是頁
回口口
霍霍
石石
確石霍
相木目
圣又土
経糸圣
以以
央央
決決
ヒヒ
旨ヒ日
指才旨
卜卜
占卜口
点占
犬犬
斗斗
状斗犬
另另
別另
不己
己言己
七七
切七刀
活舌
聞門耳
門門
同門一口

首首
道言
信寸日
導イ導
得イ士
仕込入
込ナ
ナ有
直十目
置直
安ハ女
他イ也
来来
父父
孝孝
教父
周父
調父
佳佳
集佳
利佳
兄兄
兑兑
説兑
付イ寸
具一寸
真一寸
廿一寸
广一寸
度广
加广
已已
走走
起走
必必
進進
世世

Liite 2: Muokattu opiskelujärjestys

Tutkielmassa luotu muokattu opiskelujärjestys. Harmaalla on merkitty jooyoo-kanjien ulkopuolelta olevat komponentit. ”Nostettu” tarkoittaa, että merkki on nostettu kyseiseen osioon pidemmältä opiskelujärjestystä, ja ”laskettu”, että merkki on siirretty myöhemmin opiskeltavaksi. Suuntaa antavat käännökset generoitu automaattisesti KanjiDic2-tietokannasta, mutta joitakin on lyhennetty, jotta ne mahtuvat yhdelle riville.

1. Ensimmäiset merkit, aikoja		3. Numeroita, sijainteja	
一 一	one	四 四 (nostettu)	four
二 二 (nostettu)	two	五 五 (nostettu)	five
三 三 (nostettu)	three	六 六 (nostettu)	six
刀 刀	sword	七 七 (nostettu)	seven
八 八	eight	九 九 (nostettu)	nine
分 八刀	part	十 十 (nostettu)	ten
人 人	person	勿 勿	not
日 日	day	易 勿日一	open
門 門	gate	場 土易	location
間 門日	interval	へ へ (nostettu)	ei käännöstä
木 木	tree	口 口 (nostettu)	mouth
本 木一	book	合 へ一口 (nost.)	fit
寸 寸	measurement	上 上	above
土 土	soil	下 下 (nostettu)	below
寺 土寸	Buddhist temple	中 中 (nostettu)	in
時 日寺	time		

2. Viikonpäivät		4. Koulumaailma 1	
月 月 (nostettu)	month	𠂇 𠂇 (nostettu)	ei käännöstä
火 火 (nostettu)	fire	子 子 (nostettu)	child
水 水 (nostettu)	water	学 𠂇子(nostettu)	study
金 金 (nostettu)	gold	生 生 (nostettu)	life
丁 丁 (laskettu)	street	𠂇 𠂇 (nostettu)	kettle lid radical
𠂇 一丁 (laskettu)	to take small steps	𠂇 𠂇 (nostettu)	upside-down box
𠂇 𠂇 (laskettu)	stop	高 𠂇口𠂇 (nostettu)	tall
行 𠂇𠂇 (laskettu)	going	几 几 (nostettu)	legs radical
大 大	large	𠂇 𠂇(nostettu)	ei käännöstä
小 小 (nostettu)	little	先 𠂇几 (nostettu)	before
心 心	heart	× ×	ei käännöstä
田 田	rice field	𠂇 𠂇	ei käännöstä
思 田心	think	𠂇 𠂇	ei käännöstä
		气 𠂇一𠂇	spirit
		氣 气×	spirit

5. Sekalaisia 1

夫 夫	ei käännöstä
𠂇 𠂇	shaped crown
実 𠂇夫	radical 125
𠂇 𠂇 (laskettu)	someone
者 𠂇日 (laskettu)	ei käännöstä
(nostettu)	now
今 𠂇 (nostettu)	bright
明 日月(nostettu)	exit
出 出	enter
入 入 (nostettu)	ri
里 田土	king
王 王	logic
理 王里	snout
𠂇 𠂇	ei käännöstä
𠂇 𠂇	hit
当 𠂇 𠂇	

6. Arkimaailma 1

豕 豕 (nostettu)	pig
家 𠂇豕 (nostettu)	house
母 母 (nostettu)	mother
父 父(nostettu)	father
交 𠂇父 (nostettu)	mingle
校 木交 (nostettu)	exam
𠂇 𠂇	ei käännöstä
立 立	stand up
音 立口	to spit out
部 音 𠂇	section
力 力	power
男 田力 (nostettu)	male
女 女 (nostettu)	woman
重 重	heavy
動 重力	move
イ イ (laskettu)	radical number 9
体 イ本 (laskettu)	body

7. Työelämä

𠂇 𠂇	I
云 二𠂇	say
会 𠂇云	meeting
𠂇 示 (nostettu)	cult
社 𠂇土 (nostettu)	company
仕 イ土 (nostettu)	attend
事 事 (nostettu)	matter
文 文	sentence
对 文寸	vis-a-vis
必 必 (nostettu)	invariably
𠂇 𠂇	radical 146
要 𠂇女	need
目 目 (nostettu)	eye
又 又	again
耳 耳	ear
取 耳又	take
最 日取	utmost

8. Koulumaailma 2

音 音	sound
意 音心	idea
未 未 (nostettu)	un-
味 口未 (nostettu)	flavor
貝 目八 (nostettu)	shellfish
斤 斤 (nostettu)	axe
質 貝斤 (nostettu)	substance
問 門口 (nostettu)	question
頁 頁 (nostettu)	page
是 日正 (nostettu)	just so
題 是頁 (nostettu)	topic
𠂇 𠂇	ei käännöstä
可 𠂇口	can
何 イ可	what
語 言五口 (nostettu)	word

9. Suuntia ja määriä

ナ ナ (nostettu)	ei käännöstä
右 ナ口 (nostettu)	right
工 工 (nostettu)	craft
左 ナ工 (nostettu)	left
手 手	hand
持 持寺	hold
用 用	utilize
マ マ	ei käännöstä
甬 マ用	road with walls
通 通	walk
通 通甬	traffic
手 手	hand
夕 夕	evening
多 夕夕	many
少 小 丿 (nostettu)	few

10. Arkimaailma 2

舌 舌 (nostettu)	tongue
氵 氵 (nostettu)	water
活 氵舌 (nostettu)	lively
言 言 (nostettu)	say
話 言舌 (nostettu)	tale
天 一大	heavens
ㄣ ㄣ	ei käännöstä
关 ㄣ天	frontier pass
関 門关	connection
系 丿系 (nostettu)	lineage
係 イ系 (nostettu)	person in charge
見 見	see
現 王見	present
在 ナ 丨 土 (nostettu)	exist
全 へ王	whole
艮 艮 (nostettu)	northeast
丶 丶 (nostettu)	dot
食 艮 へ 丶 (nostettu)	eat

11. Liikettä

来 来 (nostettu)	come
帚 ヅ 巾 (nost.)	broom
帰 丨 帚 (nostettu)	homecoming
玉 丶 王 (nostettu)	jewel
口 口 (nostettu)	box
国 口 玉 (nostettu)	country
戸 戸 (nostettu)	door
所 戸 斤 (nostettu)	place
也 也 (nostettu)	to be (classical)
地 土 也 (nostettu)	ground
卜 卜 (nostettu)	divining
外 夕 卜 (nostettu)	outside
丁 丁	ei käännöstä
戌 戌	5th calendar sign
成 戌 丁	turn into
牛 牛	cow
物 牛 勿	thing
方 方	direction

12. Tietämistä, tuntemista

丂 丂 (laskettu)	breath (qi)
考 𠂆 丂 (laskettu)	consider
疋 疋	ei käännöstä
定 ㄣ 疋	determine
矢 丿 天	dart
知 矢 口	know
自 自	oneself
忄 忄	heart
性 忄 生	sex
乍 乍	though
作 イ 乍	make
弋 弋	piling
代 イ 弋	substitute
七 七	ei käännöstä
化 イ 七	change
禾 禾	two-branch tree
私 禾 厶	private

13. Värejä

赤 赤 (nostettu)	red
火 火 (nostettu)	fire
黒 田土 火 (nostettu)	black
青 丰月 (nostettu)	blue
情 忄青 (nostettu)	feelings
羊 羊 (nostettu)	ei käännöstä
幸 土羊 (nostettu)	happiness
報 幸艮 (nostettu)	report
年 年	year
久 久	late
亦 亦	also
変 亦久	unusual
衣 衣	ei käännöstä
丰 丰	ei käännöstä
表 丰衣	surface

14. Koulumaailma 3

去 土ム	gone
法 讠去	method
聿 聿	brush
書 聿日	write
水 水	ei käännöstä
亲 立水	ei käännöstä
新 亲斤	new
欠 欠	lack
彳 彳	water radical
次 彳欠	next
士 士	gentleman
吉 土口	good luck
糸 糸	thread
結 糸吉	tie
果 田木 (nostettu)	fruit
止 止 (nostettu)	stop
正 一止 (nostettu)	correct

15. Lisää ajanilmauksia

么 么 (nostettu)	short thread radical
後 彳久么 (nostettu)	behind
名 夕口	name
皮 皮	pelt
彼 彳皮	he
咸 口戌一	all
感 咸心	emotion
無 無	ei käännöstä
無 無 火	nothingness
近 辶斤	near
内 内	inside
刂 刂	knife
艹 艹	grass
前 艹月刂	in front

16. Sekalaisia 2

相 木目 (nostettu)	inter-
能 亠月匕 (nostettu)	ability
葉 世木 (nostettu)	table
艹 艹 (nostettu)	grass
葉 艹葉 (nostettu)	leaf
吏 吏	officer
使 彳吏	use
車 車	car
特 牛寺	special
另 另 (nostettu)	ei käännöstä
別 另刂 (nostettu) 勹	separate
勹 (laskettu)	wrapping enclosure
勺 勺、 (laskettu)	ladle
丿 丿 (laskettu)	katakana no radical
白 丿日 (laskettu)	white
的 白勺 (laskettu)	bull's eye

Liite 3: Kanjien komponenttilistaan tehdyt muutokset

Tein Chise-projektin komponenttitietokantaan lukuisia muutoksia, jotta se vastaisi paremmin Loachin ja Wangin (2016) algoritmin periaatteita. Lisäksi muutosten myötä jooyoo-merkkien ulkopuolelta opiskeltavien komponenttien määrä pieneni 681:stä 503:en.

Radikaalien varianttien yhdistäminen

Yhdistin tavanomaiset radikaalien variantit alkuperäiseen merkkiin, sillä Loachin ja Wangin (2016) algoritmi määrittelee varianttien oppimisen hinnaksi 1 (+ alkuperäisen merkin oppimisen hinta). Muutoksia yhteensä 9 kappaletta.

Esimerkiksi 𠂇 yhdistetty alkuperäiseen merkkiin 示 'alttari'. Myös radikaalit 月 'liha' ja 月'kuu' on yhdistetty, sillä nykykielessä ne näyttävät samalta.

Monimutkaisten ja harvoin esiintyvien komponenttien jakaminen alakomponentteikseen

Suurelta osin poistin komponentteja, joita ei löydy merkistöstandardeista (Chine-tietokannassa &CDP-89AB;-tyylisillä koodeilla) ja korvasin ne yleisemmillä alakomponenteilla. Useimmat poistetuista komponenteista esiintyivät vain yhdessä merkissä, jos ne esiintyivät useammassa, tein saman muutoksen kaikkiin.

Muutoksia tein yhteensä 127 kappaletta. Näitä muutoksia oli määrällisesti eniten, mutta ne lyhensivät ja järkeistivät huomattavasti opiskeltavien komponenttien määrää.

Esimerkiksi sekvenssissä

&CDP-8BB5; &CDP-8BB5;
然 &CDP-8BB5; 𠂇

Komponentti &CDP-8BB5; jaettiin kahtia:

然 夕 犬 𠂇

Turhan komponenttijaon poistaminen

Poistin komponenttijaon joissakin tapauksissa, mikäli poistaminen tuotti selkeämmän lopputuloksen. Muutoksia yhteensä 30 kappaletta.

Esimerkiksi:

出 𠂇 𠂇 → 出 出

Monet näistäkin muutoksista auttoivat vähentämään harvinaisten ja turhien komponenttien opiskelua:

&CDP-8DBF; &CDP-8DBF;

麤 &CDP-8D56;&CDP-8DBF; → 麤 麤

Lähes identtisten komponenttien sulauttaminen

Radikaalien varianttien lisäksi sulautin toisiinsa muita olennaisilta osin identtisiä komponentteja. Muutoksia yhteensä 8 kappaletta.

Esimerkiksi komponentit 𠂇 ja 𠂈 sulautettiin.

Moninkertaisten komponenttien poistaminen

Koska Loachin ja Wangin (2016) algoritmin mukaan komponenttien toistumista ei huomioida vaan kukin komponentti lasketaan vain kerran, poistin tietokannasta ylimääräiset komponentit.

Muutoksia yhteensä 10 kappaletta.

Esimerkiksi: 高 𠂇𠂇𠂇𠂇 → 高 𠂇𠂇𠂇

Muut muutokset

Vaikeasti luokiteltavia muutoksia tein yhteensä 19 kappaletta. Joissain tapauksissa tein useammantyyppisiä muutoksia samaan komponenttiin/merkkiin, esimerkiksi:

𠂇 𠂇

替 𠂇日

→ 替 夫日 (turhan välikomponentin poistaminen ja moninkertaisuuden poistaminen)

Joitakin tietokannassa jakamattomana kokonaisuutena pidettyjä yleisiä merkkejä jaoin komponentteihin, esimerkiksi

果 果 → 果 田木

Muutamia muita lajittelemattomia muutoksia olivat esimerkiksi komponenttijaon muuttaminen toiseksi:

易 旦勿 → 易 勿日一

Liite 4: Algoritmin ohjelmakoodi

Kaikki projektiin liittyvä koodi ja data löytyy internetistä osoitteesta:

<https://github.com/ahaavisto/ahaavisto.github.io/tree/master/gradukoodi>

Varsinainen algoritmin koodi on lyhyt.

```
def algo():
    global jarjestettavat_merkki
    for merkki in jarjestettavat_merkki:
        algon_rekursio(merkki, merkin_indeksi)
```

[illegible]

Liite 5: Kuvaaja opiskeltujen merkkien suhteesta korpuksen yleisimpiin merkkeihin

Opiskeltujen merkkien osuus korpuksen yleisimmistä 250 merkistä, kun algoritmin osalta on huomioitu ainoastaan opitut jooyoo-kanjit eikä niitä varten opiskeltuja komponentteja. Luvussa 5.3. on kuvaajasta versio, jossa mukana on huomioitu myös komponentit.

